THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 9, November 1977
Printed in U.S.A.

Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide—I

By DAVID J. THOMSON

(Manuscript received April 7, 1977)

Techniques for reliably estimating the power spectral density function for both small and large samples of a stationary stochastic process are described. These techniques have been particularly successful in cases where the range of the spectrum is large. The methods are resistant to a moderate amount of contaminated or erroneous data and are well suited for use with auxiliary tests for stationarity and normality. Part I is concerned with background and theoretical considerations while examples from the development and analysis of the WT4 waveguide medium will be discussed in Part II, next issue.

I. INTRODUCTION

The problem of estimating the spectrum of a stationary time series has appeared frequently in the scientific literature and myriad approaches have been suggested. Nonetheless it became apparent during the course of the development of the WT4 waveguide system that these methods were inadequate for many of the data sets of interest. The techniques presented here were therefore developed.

It is commonly stated that the method selected to estimate a spectrum depends on the ultimate use of the estimate, and unfortunately to some extent this is true. The method described below is felt to represent an advance in that the basic technique works well in a variety of cases which previously would have required individual treatment. The loss calculations reported in Anderson *et al.*[1] are indicative of its accuracy.

The procedure which has evolved for estimating spectra can best be described as robust adaptive prewhitening. Such methods have three distinct stages: formation of a *pilot* spectrum estimate, using this estimate to design a *prewhitening* filter, and finally giving the result as the *ratio* of the spectrum of the filtered data to the power transfer function of the filter. This method is potentially both efficient and robust. The

efficiency of a statistical estimation procedure is the fraction of the information, in the sense of Fisher,[2] conveyed hy the estimate ahout the parameter heing estimated to the total information on this parameter inherent in the data. An estimation procedure is *robust* if it remains efficient over a wide range of conditions and is relatively immune to a small fraction of outlying or erroneous data.

For the sequential method descrihed here to he efficient, the pilot estimate must he designed to have a large dynamic range at the expense of frequency resolution. The second spectrum estimate, which works on the filtered data, uses the opposite choice and so is chosen on the hasis of frequency resolution. This can he done without incurring a large penalty in loss of effective dynamic range as this information, acquired by the pilot estimate, has heen transferred to the filter specification. In one meaning of the term this procedure is rohust in that it can normally handle situations where either estimate alone would fail. By using a nonlinear filter for the prewhitening operation the procedure may also he made rohust in the sense that it is resistant to moderate amounts of erroneous or contaminated data.

In this method the pilot estimate of spectra is a comhination of several direct estimates of spectra computed on suhsets of the data using a window defined by a prolate spheroidal wave function. Using this estimate as a basis an autoregressive model of the process is formed. This model is then used to generate a *nonlinear* prediction error filter. The output of this filter consists of prediction residuals from a modified data sequence and is quite immune to occasional isolated errors in the data.

Section II gives an overview of the complete estimation procedure so that the descriptions of the individual stages of the process are taken in the proper perspective. Section III is a review of properties of direct estimates of spectra which are used for hoth the pilot and final estimation procedures. Sections IV to VIII descrihe the several stages of the estimation procedure in detail. While these sections contain some examples they are primarily concerned with theory and hackground. Part II will consist primarily of examples and comparisons with standard techniques.

It should he emphasized that the same approach is used for *both* short and long data sets and that the only difference hetween these cases is one of detail and not philosophy. We define a *short* time series as one which cannot he suhdivided into suhsets having almost uncorrelated spectrum estimates.

Since this technique is hasically nonparametric, it is frequently asked whether a parameterized estimate of spectrum might not give hetter results. It has been shown hy Arato[3] that *only* for the autoregressive case can a process he descrihed hy a fixed numher of sufficient statistics and

that in general the number of sufficient statistics increases with the sample size. As a result, efficient parametric estimates are not likely to be even conceptually simpler than the nonparametric estimates used here.

It is also asked why maximum-likelihood techniques are not used directly, and, while asymptotic results on parametric maximum likelihood estimates of spectra are available in Whittle,[4] constructive procedures for obtaining nonparametric maximum-likelihood estimates of the spectrum of a stationary Gaussian time series are unknown. It is, however, possible to check if a *given* estimate is maximum likelihood or not. This test, described in Thomson,[5] depends on the Karhunen-Loève expansion of a random process (see Loève[6]). In this test the data is expanded in terms of the sample eigenfunctions of the spectrum estimate, and, if this estimate is maximum likelihood, the expansion coefficients, $\hat{a}_n$, will satisfy the conditions $\hat{a}_n^2 = \hat{\lambda}_n$ in which $\hat{\lambda}_n$ are tbe corresponding sample eigenvalues. By the Szegö theorem (see Grenander and Szegö[7]) this comparison is asymptotically equivalent to comparisons on the spectrum at a frequency spacing of $1/T$. This agrees with the conventional Rayleigh resolution and heuristically a spectrum estimate with this resolution and low bias is likely to be efficient. Tbis argument provides the motivation for the present technique. Simple data windows with frequency resolution close to $1/T$ do not provide enough bias protection. Moreover this is not just a result of not having chosen the right "simple" data window but the result of fundamental characteristics of the Fourier transform (see Landau and Pollak[8]). Data windows like the 4π prolate spheroidal wave function which provide the protection from bias have frequency resolution on the order of $4/T$ and so are inefficient from this viewpoint. It must be emphasized that the sequential approach used here potentially has both limitations since it cannot resolve details spaced by $1/T$ in frequency when their levels are more than 4 or 5 decades apart. On the other hand if the spectrum is not quite so pathological and varies "slowly" over 10 to 15 decades then the method can provide frequency resolutions approaching $1/T$ with relatively low bias.

II. SUMMARY OF THE ESTIMATION PROCEDURE

2.1 Data preparation

At the beginning the data is plotted, and serious outliers, missing values, etc., edited by use of either interpolation or successive prediction and interpolation. These predictors and interpolators are the optimum linear forms based on previous spectra estimates of a similar process or on assumed valid data from the current sample. It is also frequently necessary to remove the mean value function of the "cleaned" data. This

is always done in the analysis of individual tubes to eliminate the curvature resulting from gravitational sag.

2.2 Pilot spectrum estimate

For the remainder of this paper we assume that the available data is a sequence of samples $\{x_t\}$, $t = 0, 1 \cdots L$, and that the sampling interval has been normalized to 1. Consequently the normalized Nyquist frequency is $\frac{1}{2}$. Both because the notation is more compact and also as a reminder that the basic processes are continuous† most operations will be denoted by integrals. In the actual computations most of these integrals are replaced by simple sums but on occasion spline approximations to the integrals (see Aronson[10]) are used. The frequency variable will be denoted by f with $\omega = 2\pi f$.

The initial spectrum estimate is normally computed using a variation of Welch's[11] method: the basic data set is divided into k overlapping *subsets* each of length T and offset from the previous one by a distance b. The data from each subset is tapered using a zero order *prolate spheroidal wave function*, with parameter $c = 4\pi$ and the Fourier transform of the result computed. The raw estimate of spectra on the jth subset, $\hat{S}_j(\omega)$, is then the squared magnitude of the transform so that its univariate distribution is proportional to a χ^2_2. The use of the prolate data window guarantees, under simple conditions, that the bias of the estimate within each subset is of purely *local* origin and that estimates separated by more than $2c/T$ in frequency are essentially uncorrelated. However, to account for the correlation induced by the tapering the total number of degrees of freedom must be reduced. These effects and the bivariate distribution of the estimates is discussed in Section 3.2.

Because the raw estimates, $\hat{S}_j(\omega)$, are very volatile it is often desirable to smooth the different subset estimates. These estimates, smoothed to have ν degrees of freedom, will be denoted by $\bar{S}_j(\omega)$. In the original Welch technique the pilot estimate of spectrum, $\bar{S}(\omega)$, is the arithmetic average of the subset estimates. When the data contains outliers it is advantageous to replace the simple average with a robust combination of the subset estimates as discussed in Section V. Both because it is based on subsets of the data and because of the heavy tapering implied by the use of the parameter $c = 4\pi$ (see Section III) the pilot spectrum estimate has poor frequency resolution compared to the final estimate of spectra. For reasons discussed below excessive resolution in the pilot estimate is frequently counterproductive and this technique produces a stable estimate with adequate bias protection in situations where the range of the spectrum is very large.

† The paper by Dzhaparidze and Yaglom[9] contains information on the complexities induced by sampling basically continuous records.

2.3 Tests for stationarity

Large data sets can be tested for stationarity using the method described in Thomson.[12] Briefly the approach compares the different subset estimates, $\overline{S}_j(\omega)$, using Bartlett's M statistic for heteroscedasticity of variance between subsets at constant frequency. Equally spaced samples of the test statistic, $M(\omega_j)$, are then pooled and tested for conformance to the distribution expected for homogeneous samples.

2.4 Construction of autoregressive models

Stationary time series have four generally accepted canonical representations; Cramér's orthogonal increment spectral representation, the Karhunen-Loève expansion, the moving average, and autoregressive models. Of these the autoregressive model is perhaps the most useful for making inferences on the structure of the process. For further information see the review paper by Kailath.[13]

Most autoregressive methods either begin with a sample autocorrelation function and solve the Yule-Walker equations directly (Makhoul[14]) or else resort to a variation of Wiener spectral factorization (Whittle[15]) applied to an estimate of spectra; neither approach is entirely satisfactory. For the estimation of waveguide spectra both methods have been used and in Section VI a method combining the better features of both is discussed. In cases where the range[‡] of the spectrum is relatively small, solving of the Yule-Walker equations using Durbin's modification of the Levinson algorithm (see Section VI) is satisfactory. In this case the autocorrelations used are obtained by Fourier-transforming the pilot estimate of spectra. When the range of the spectrum is larger tbe Wiener technique is more stable but results in a very long predictor. Backward application of the Levinson algorithm may then be used to generate a more compact representation. In both cases the order p of the autoregressive representation has usually been chosen on the basis of Parzen's[16] stopping rule and the innovations variance corresponding to the pilot spectrum $\overline{S}(\omega)$. Details of the procedure are given in Section VI.

The autoregressive representation has an intuitive explanation in waveguide applications in which the prediction can be thought of as analogous to a local "warped normal mode" representation and the innovations process the changes required in the field configuration to maintain the "local" character. The casual nature of the autoregressive representation corresponds to propagation in the forward direction so that the field configuration at a given point reflects distortions which have been passed but not those in the future.

[‡] The *range* of a spectrum refers to the *logarithmic* range or the ratio $\max\{S\}/\min\{S\}$.

2.5 Prewhitening, robust filtering

The autoregressive model formulation gives the casual filter which, for fixed impulse duration, p, has minimum output power. The residual sequence or output of such a filter (known as a *prediction error filter*) is the difference hetween the ohserved and predicted values of a data sequence *using the previous p data points as a base for the prediction.* When the autoregressive model is correct the residual sequence will be serially uncorrelated and have a white spectrum. When the data contains outliers the effect of such filtering is to contaminate the p residuals following each erroneous point.

The *robust filter algorithm* is a nonlinear procedure based on an autoregressive model which is designed to reduce the effects of occasional outliers. The output of this filter or the *modified data sequence* is an estimate of the uncontaminated process. This sequence is formed hy comparing successive input data points with the value predicted from the modified sequence. In regions where the prediction errors are "small" relative to the innovations variance, the modified sequence is essentially a copy of the input data. When the prediction errors are "large," the corresponding points of the modified data sequence are the predictions rather than the data and for intermediate prediction errors the behavior depends on a weight function. When the modified data sequence is used as a hasis for the final estimate of spectra, the prediction error sequence is the difference hetween the predictions and the value of the modified data sequence. For uncontaminated data this corresponds to the output of the linear prediction error filter hut when a large error is present the algorithm has two effects: first, the large output residual is replaced hy a zero; second, hecause of the feedhack nature of the method, propagation of the error into suhsequent predictions is greatly reduced. As with all methods which alter or ignore extreme ohservations a compromise must he drawn between rejecting some valid data and accepting occasional errors and, in the robust filter algorithm, this compromise is reflected in the choice of weight function. In Section VII a weight function motivated by the normal extreme value distrihution which has hoth intuitive appeal and desirable mathematical properties is descrihed.

2.6 Final estimate of spectrum

The prediction residuals, or output of the prediction error filter, are the sequence which has minimum power for a filter whose impulse response has duration p. Consequently in the frequency domain the effect of such an operation is necessarily to reduce the highest parts of the spectrum first. As the complexity of the filter is increased the residual spectrum approaches a constant at which point further improvement is impossihle. In practive finite order autoregressive filters seldom attain

this limit but rather have the effect of reducing the range of the spectrum, usually without following any fine structure which is present and, as a result, information describing the fine structure is left in the residual process. On occasions when the autoregressive fit is forced to follow too fine structure in the spectrum the spectrum of the residuals may be locally more complex than the spectrum of the original process.†

Since the range of the spectrum has been reduced the procedure used to estimate the spectrum of the residuals is designed to have high-frequency resolution at the expense of sidelobe suppression.

When the nonlinear version of the prediction error filter is used it is commonly observed that the pilot spectrum, estimated from the contaminated data, is considerably higher than the final estimated spectrum at frequencies where the spectrum is small. So that these differences are not obscured with bias the pilot final taper must be such that the corresponding spectral window decays significantly with frequency and consequently tapers such as the Taylor equiripple design (see Rife and Vincent[17]) are inadvisable. The window which has been used most for this purpose is Tukey's spliced cosine taper. For long data sets this window is satisfactory but with very short sets, for example individual waveguide tubes, the first sidelobe of this window is too high and a more complex window described by a series expansion in prolate spheroidal wave functions is used.

The final estimate of spectrum is based on an approximation introduced in Grenander and Rosenblatt,[18] which is that the predictor and prediction residuals are statistically independent. Under this assumption the final estimate of spectrum will be the spectrum of the residuals divided by the power transfer function of the prediction error filter.

2.7 Smoothing

One of the most commonly recommended operations in spectrum estimation is that of smoothing the raw estimates by means of local averaging over frequency. Contrary to these recommendations the final estimates of spectra are almost never smoothed. Moreover, in cases where "smoothed" estimates of spectra are used, the smoothing is frequently the result of nonlinear and adaptive procedures. Such smoothing is useful in plotting applications, and for improving the stability of pilot spectrum estimates from short data sets. Certain nonlinear smoothers are also very useful for finding low level lines in complex spectra.

The general philosophy of these methods has been to test the raw spectrum for local homogeneity: when the local spectrum appears to be

† For spectrum estimation problems a good measure of complexity is $\left|\frac{\partial^2 S(\omega)}{\partial \omega^2}\right| \Big/ S(\omega)$.

homogeneous it is smoothed, but in cases where the raw spectrum exhibits variations greater than normal, a typical response is to reduce the width of the smoother. A second approach which is used is to initially "smooth" the raw spectrum using a robust nonsymmetric location estimate and then to put the peaks back on the basis of "inverse influence."

III. DIRECT ESTIMATES OF SPECTRA

In both the pilot and final phases, the spectrum is estimated by the so-called *direct method* and while the parameters and application of the estimator are different in the two cases, the basic form is the same. Information on direct estimates is available from several sources, for example Blackman and Tukey,[19] Jones,[20–22] Tukey,[23] Koopmans,[24] Brillinger.[25] In this section properties of the direct estimate are reviewed and compared to the indirect estimate; the role of prolate spheroidal wave functions as a means of reducing the bias of the estimate is described and compared to standard data windows. The next subsection describes the variance of the estimates with emphasis on characteristics of prolate windows and smoothing when the estimates included in the smoother are correlated. The final subsection is concerned with Welch estimates and a technique for choosing the optimum subset spacing.

The *direct* estimate of spectrum is defined by

$$\hat{S}_D(\omega) = \left| \int_0^T e^{i\omega t} D(t) x(t) dt \right|^2 \tag{1}$$

In this definition the data, x, is defined on the domain $[0, T]$, ω is radian frequency, and D is a *data window* or *taper*. The data window is normalized according to the convention

$$\int_0^T D^2(t) dt = 1 \tag{2}$$

so that the resulting spectrum is interpretable in physical units.

Almost all of the published estimates of spectra are either direct estimates, smoothed direct estimates, or rational fits to direct estimates. When D is constant $\hat{S}_D$ is the periodogram. Smoothing the *extended* periodogram† with appropriate weights corresponds to the various *indirect* estimates. Similarly an autoregressive or "maximum entropy" estimate may be regarded as an all-pole rational fit to the extended periodogram and Pisarenko[26] estimates constitute a generalization of

† In the simple periodogram estimates are computed at a frequency spacing of $1/T$ and the corresponding autocorrelations are circularly defined. A frequency spacing $<1/2T$ is used in the extended periodogram and its Fourier transform yields the common autocorrelations.

this idea.[‡] The notable exceptions are the Whittaker periodogram and the Burg estimate (see Section 6.4).

The application of smoothers or curve-fitting procedures to the basic estimate conceals its true nature and the fact that the properties of these estimates are controlled primarily by the data window D. For example it is commonly stated that the fundamental uncertainty in spectrum estimation is between resolution and variance and more papers than it is convenient to list have worked on better "lag windows" to minimize this conflict. Unfortunately the emphasis on this secondary problem has masked the primary uncertainty between resolution and bias. The basic problem with indirect estimates and the lag window approach is that it represents an attempt to patch the periodogram. A more logical approach is to start with a better basic spectrum estimate.

Despite its simplicity the direct estimate is not well understood. In particular the differences between direct estimates using *data windows* and indirect estimates using *lag windows* are frequently confused.

The expected value of the direct estimate (1) may be written

$$\mathrm{E}\{\hat{S}_D(\omega)\} = \int_0^T \int_0^T e^{i\omega(t-u)} D(t) D(u) \mathrm{E}\{x(t)x(u)\}\, dt\, du \tag{3}$$

For second order or covariance stationary processes the autocovariance function is defined by

$$R(\tau) = \mathrm{E}\{x(t)x(t+\tau)\} \tag{4}$$

and may be represented in terms of the spectral density function by using the Wiener-Khintchine relation

$$R(\tau) = \frac{1}{2\pi} \int e^{i\omega\tau} S(\omega)\, d\omega \tag{5}$$

and denoting the Fourier transform of the data window D by $\tilde{D}$ one obtains

$$\mathrm{E}\{\hat{S}_D(\omega)\} = S(\omega) * |\tilde{D}(\omega)|^2 \tag{6}$$

where S is the true spectrum of the process and * indicates convolution. Since D is a *time-limited* function $\tilde{D}$ is an entire function of ω so that the direct estimate is biased for all spectra which are not white. The function $|\tilde{D}(\omega)|^2$ is known as the *spectral window* of the estimate.

An alternative description results from expressing eq. (3) in terms of the autocovariance function, R, of the process as

$$\mathrm{E}\{\hat{S}_D(\omega)\} = \int_{-T}^{T} e^{-i\omega\tau} L_D(\tau) R(\tau)\, d\tau \tag{7}$$

[‡] The Capon[27] estimate, while superficially similar, is intended for estimating the magnitude of periodic components in a background of a known covariance structure.

where the convolution D^*D has been identified as an "equivalent" *lag window*, $L_D(\tau)$. Because of this identification characteristics of the *indirect* estimate of spectra,

$$\hat{S}_L(\omega) = \int_{-T}^{T} e^{-i\omega\tau} L_D(\tau) \hat{R}_u(\tau)\, d\tau \tag{8}$$

using an unbiased estimate of autocovariance

$$\hat{R}_u(\tau) = \frac{1}{T - |\tau|} \int_0^{T-|\tau|} x(t)x(t + |\tau|)\, dt \tag{9}$$

are often used incorrectly to describe the *direct* estimate, $\hat{S}_D(\omega)$. Except for their first moments these two estimates have few properties in common: one very important difference is that *the direct estimate is positive* while the "equivalent" indirect form need not be. Also, because their common spectral window enters the estimate in fundamentally different ways, the variances of the two estimates are different.

3.1 Minimum bias estimates and prolate spheroidal wave functions

The most convenient description of bias induced by the data window is through the spectral window $|\tilde{D}|^2$ as expressed in eq. (6). The effect of this convolution is to change the apparent distribution of power in a complex manner and, since all windows cause some redistribution, a minimal requirement is that the indicated power be left "close" to its original location. Defining "close" to be within a tolerance Ω of ω we require that the *broadband bias*, i.e., bias from outside $(\omega - \Omega, \omega + \Omega)$, be small. Denoting this bias by $B_B(\omega)$ and the integral over frequency with the section $(\omega - \Omega, \omega + \Omega)$ excluded by $\oint$ we have

$$B_B(\omega) = \frac{1}{2\pi} \oint S(\omega - \zeta)|\tilde{D}(\zeta)|^2\, d\zeta \tag{10}$$

From the definition of a direct estimate and the convolution theorem the broadband bias in a particular sample is

$$\hat{B}_B(\omega) = \left| \oint \tilde{x}(\omega - \zeta)\tilde{D}(\zeta)\, d\zeta \right|^2 \tag{11}$$

where $\tilde{x}(\omega)$ is the spectral representation of x (see Doob,[28] chapter 10). By the Cauchy inequality this bias may be bounded so that

$$\hat{B}_B(\omega) \leqslant \frac{1}{2\pi} \oint |\tilde{x}(\omega - \zeta)|^2\, d\zeta \frac{1}{2\pi} \oint |\tilde{D}(\omega - \zeta)|^2\, d\zeta \tag{12}$$

The first factor of this inequality depends only on the process and, as the integrand is positive, is simply bounded by adding the integral from $\omega - \Omega$ to $\omega + \Omega$ and identifying the result using Parseval's theorem. The second factor in the inequality depends only on the data window, D, and

expresses the energy in $\tilde{D}$ outside $-\Omega$, Ω. D is a time-limited function of unit energy and this inequality is minimized when D is a *prolate spheroidal wave function.* The fundamental role of these functions in relation to Fourier transforms and related problems have been described in a remarkable series of papers by Slepian and Pollak,[29] Landau and Pollak,[8,30] and Slepian.[31] When the bounds for both integrals are combined the result is that

$$\hat{B}_B(\omega) \leq \hat{\sigma}^2 T(1 - \lambda_{00}(c)) \tag{13}$$

where $\hat{\sigma}^2$ is the sample variance, $c = \Omega T/2$, and $\lambda_{00}(c)$ is the largest eigenvalue of the integral equation

$$\lambda_n \psi_n(t) = \int_{-1}^{1} \frac{\sin c(t-s)}{\pi(t-s)} \psi_n(s)\, ds \tag{14}$$

Tables of the eigenvalues of this equation have been published in Slepian and Sonnenblick[32] and asymptotic descriptions given by Slepian.[33] From the latter reference

$$1 - \lambda_{00}(c) \approx 4\sqrt{\pi c} e^{-2c} \tag{15}$$

As the width of the guard band, Ω, increases this bound decreases rapidly. For exploratory time series analysis and the formation of pilot spectrum estimates a very convenient value of c is 4π for which $1 - \lambda_{00} \approx 3 \times 10^{-10}$. In Thomson *et al.*[34] empirical studies show that direct estimates using this window are generally superior to several other spectral estimates in common use. Other examples are contained in Thomson.[5] Windows using approximations to prolate spheroidal wave functions have been described by Kaiser,[35] Eberhard,[36] and in fact the Parzen[37] window can be considered as a fourth-order successive approximation to the 4π prolate window.

Figure 1 shows the 4π prolate data window (and several other windows described below) and the low weighting given near the ends of the data are evident. The corresponding spectral windows are shown in Fig. 2, and here the reason for using the 4π prolate taper in situations where the spectrum varies over large ranges is most evident. The frequency scale of this plot has been normalized to units of $1/T$ so that by a frequency of $4/T$ the spectral window corresponding to the 4π taper has decayed by more than 10 decades. It should be noted that the curves for the other windows represent *envelopes* of the spectral windows. The actual spectral windows are similar to that shown for the compound prolate window and decay in an oscillatory manner.

When the range of the spectrum is known to be small it is clear that the use of this window is inefficient in that the frequency resolution is much less than it is for windows with higher sidelobes, and several al-

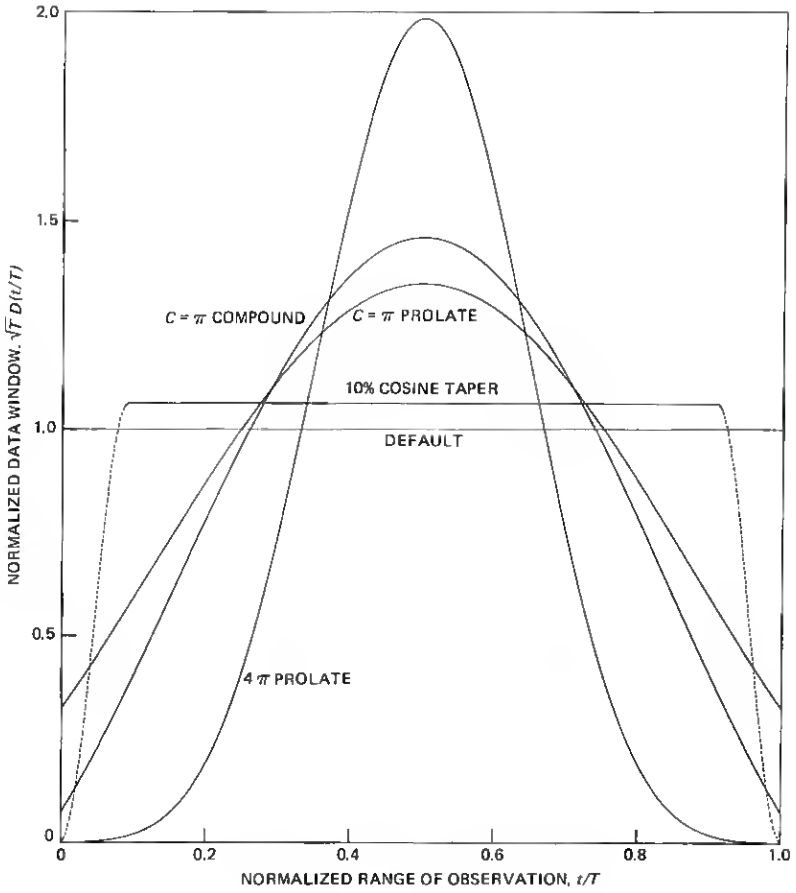

Fig. 1—Comparison of data windows.

ternatives are available: no tapering, ad-hoc tapers, and prolate tapers with lower values of c.

Very few spectra resulting from physical processes are so uninteresting that the "elimination" of tapering is ever advisable; in this case the taper actually used is $1/\sqrt{T}$ over $(0, T)$ and 0 elsewhere. This "default taper" has

$$T\left(\frac{\sin \omega T/2}{\omega T/2}\right)^2$$

as a spectral window so that, as shown in Fig. 2, the first sidelobe is only ~13 dB down from the central maxima.

Of the various ad-hoc techniques, Tukey's[23] spliced cosine taper is perhaps the most useful and it has been used for many of the final esti-

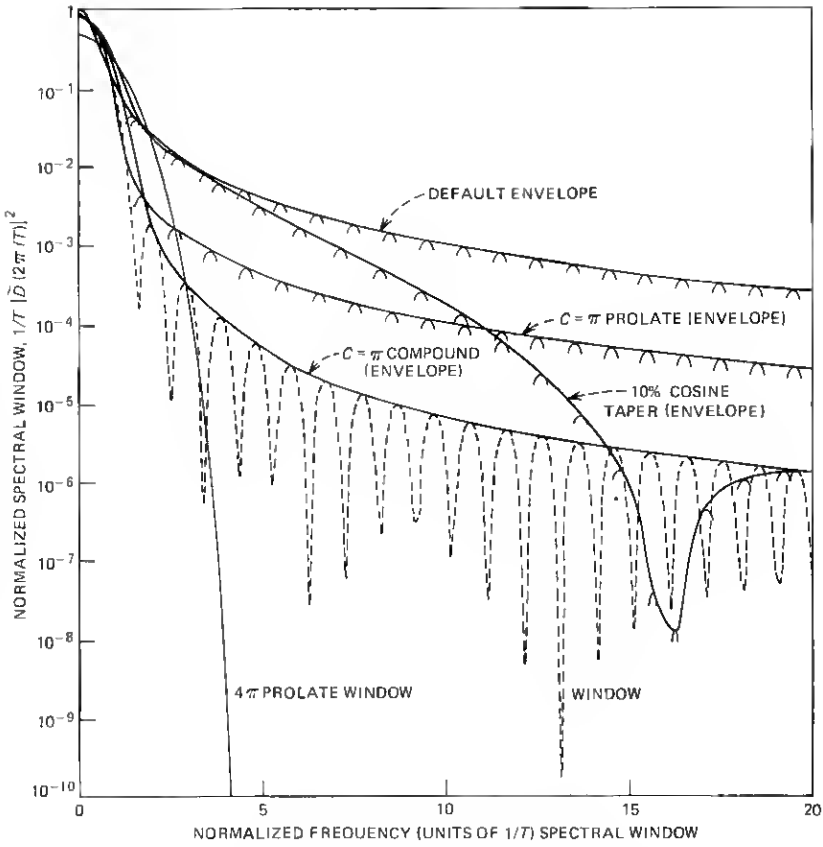

Fig. 2—Envelope of spectral windows.

mates of spectra used in the WT4 project. Since this taper, shown in Fig. 1, weights the data in a much more uniform manner the corresponding spectral window, Fig. 2, has a narrower center lobe than the 4π window and the sidelobes decay much faster than those of the default window.

Unfortunately the first few sidelobes of this window are too high for it to be usable in many applications where accurate estimates of the fine structure of a spectrum are required. This leads to considering the spheroidal wave functions again and for maximum concentration in a bandwidth $1/T$ the appropriate value of the parameter c is π. As before the maximum concentration is achieved by using the function of order 0 but for the present application a better compromise can be obtained by using a linear combination of the functions of order 0 and 2 with the coefficients determined by the additional constraint imposed by requiring that the first two sidelobes be minimized. This "compound

prolate" taper is also plotted in Fig. 1 and it is clear that the weight is less extreme than the 4π taper but distinctly different from the spliced cosine form. From the plot of the spectral windows it can be seen that the main lobe of the compound window is almost as narrow as that of the spliced cosine and also that the first sidelobes are down 27 dB instead of 13 dB. Since the widths of the main lobes are all very close to the same width, tbis gain in performance is essentially free and results from the superior characteristics of the prolate functions. It might also be mentioned that the usual objection to the use of the prolate spheroidal wave functions, namely that they are "impossible" to compute, is false and that by using Horner's rule together with the expansion given in Flammer,[38] Section 3.2, they may be computed very rapidly. Appendix A gives expansion formulae for the π and 4π prolate data windows.

In anticipation of Section VI it is also interesting to compare the bias of the estimates of autocorrelation obtained by transforming the various spectrum estimates. From eq. (7) it is apparent that such estimates of the autocorrelation function at lag τ will be biased by the factor $L_D(\tau)$. These lag windows are plotted in Fig. 3. From this figure it can be seen that the bias imposed on the low-order autocorrelations by the windowing techniques is much less than that resulting from the common positive definite estimate [obtained by replacing the factor $T - |\tau|$ in eq. (9) with T] corresponding to the simple extended periodogram. It should be noted that if this factor is divided out the resulting unbiased estimate is not positive definite and frequently results in negative "prediction variances." For fitting autoregressive models, the low-order autocorrelations are crucial and, as can be seen from the insert in Fig. 3, for $\tau/T < 0.01$ the bias obtained using the 4π prolate window is lower than that obtained from the extended periodogram on data sets *10 times as long*. The scale of such comparisons can be best judged by noting that the one-step autocorrelation in the field evaluation test curvature data is about 0.99983.

3.2 The distribution of direct spectrum estimates: liftering†

The preceding sections were adressed primarily to the problem of bias in direct spectrum estimates without particular attention being paid to their variances or distributions. Since reliable interpretation of spectrum estimates requires understanding of both their distributions and the correlations between estimates, the following sections treat these and the closely related problem of smoothing. Because of the correlations induced by the data window, the variance of smoothed direct estimates depends *both* on the smoothing weights *and* on the data window.

As mentioned in the introduction, the final estimate is rarely

† See Bogert, Healy, and Tukey[39] for definitions of these terms.

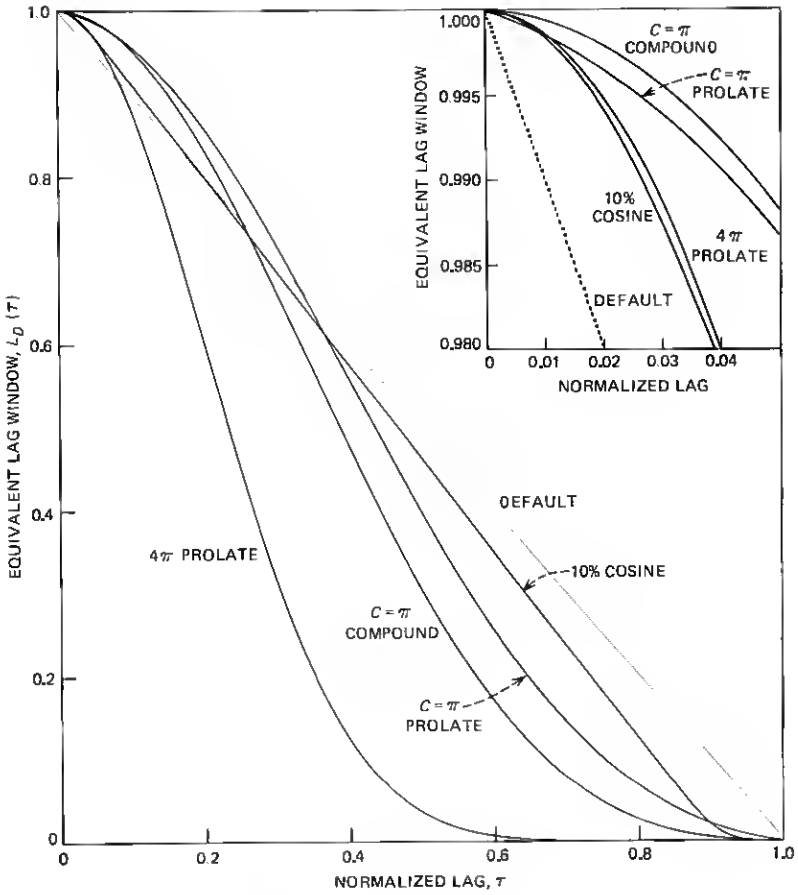

Fig. 3—Comparison of lag windows.

smoothed, but in the formation of the pilot estimate smoothing is a critical step. The primary reason for smoothing the pilot spectrum estimate is to obtain a more accurate autoregressive model. Because the direct estimate is inconsistent, that is, its first-order distribution and variance are independent of sample size, T, smoothing is imperative when spectral factorization is employed and experience has shown that serious errors are obtained when unsmoothed estimates are used with the other autoregressive modeling techniques. Also, when the robust filter algorithm is used, the prediction residuals are measured on the scale of the estimated innovations variance. The accuracy of this estimate, and hence the reliability of the procedure, depends *both* on the actual stability of the pilot estimate *and* on what we estimate that stability to be. The latter effect enters in the form of a bias correction factor, a

function of the "equivalent degrees of freedom" of the pilot estimate, on the innovations variance estimate. With "short" data sets stability can only be obtained through liftering but even with long data sets where the Welch technique is applicable, liftering is used to improve the sensitivity of the stationarity test.

3.2.1 The distribution of direct spectrum estimates

Excluding the neighborhoods of the origin and the Nyquist frequency, direct estimates of spectra are approximately distributed as a χ^2_2. For Gaussian data this result is exact and even in cases where the original data is reasonably nonnormal it is known (Fisher,[40] Bartlett[41]) to be a remarkably good approximation. Because the variance of such estimates is given by the square of their expected value, this fact emphasizes the need to start with a better estimate of spectra than the periodogram; estimates with low bias will have lower variance than estimates with high bias.

The bivariate probability density function of direct spectrum estimates can be obtained from those given by Miller *et al.*[42] for Rayleigh processes

$$p(s_1, s_2) = \frac{1}{1-\Lambda} e^{-(s_1+s_2)/(1-\Lambda)} I_0\left(\frac{2\sqrt{\Lambda s_1 s_2}}{1-\Lambda}\right) \tag{16}$$

where both s_1 and s_2 have been standardized to unit level, I_0 is the usual modified Bessel function, and Λ is the correlation between s_1 and s_2 given below by eq. (18).

The characteristics of this distribution are most easily seen by considering the conditional distribution $p(s_1|s_2)$. For this distribution a critical point is given by $s_2 = (1-\Lambda)/\Lambda$; at this point $\partial p(s_1|s_2)/\partial s_1|_{s_1=0} = 0$ which for lower values of s_2 resembles the univariate distribution and has its maximum at 0, while for larger values of s_2 the mode approaches s_2.

Figure 4 shows plots of the conditional distribution for $s_2 = 0.5$ and 1 and for values of Λ appropriate for the 4π prolate window at frequency spacings of $0.25/T$, $0.5/T$, $0.75/T$, $1/T$, and $2/T$.

3.2.2 Smoothing and frequency correlations of spectrum estimates

There is a considerable literature on smoothing spectrum estimates (see for example Blackman and Tukey,[19] Parzen,[37,43] Papoulis,[44]) and the variance and distribution of smoothed estimates (Jones,[20] Grenander *et al.*[45]) but much of this work is specialized to estimates based on the periodogram and cases where the different raw estimates included in the smoothing operation are uncorrelated. For the prolate data windows the latter assumption is unwarranted (as indeed it is even for the ex-

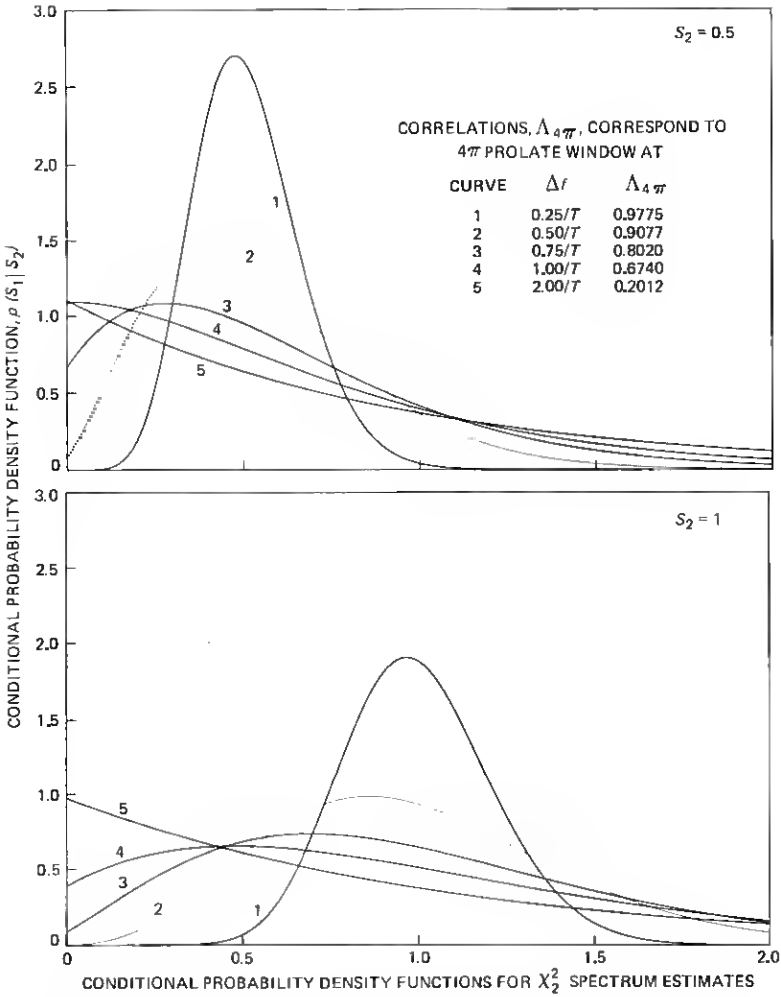

Fig. 4—Conditional probability density functions $p(S_1/S_2)$ for bivariate χ^2_2 distribution. Top set of curves, conditioning variable $S_2 = 0.5$; lower set, $S_2 = 1$.

tended periodogram) since, for the 4π window, the bias is only localized within a band of $\pm 4/T$. In the more general case where the raw spectrum estimates are correlated, smoothing over a fixed bandwidth is less effective and conventional smoothing techniques will be characterized by fewer "equivalent degrees of freedom" than given by the usual estimate. Most of the work on smoothing assumes that the true spectrum does not vary appreciably over the width of the smoother and under this approximation the influence of smoothers on direct estimates is fairly simple to evaluate.

To assess the effects of smoothing correlated spectrum estimates it

is necessary to examine their correlation properties. By expanding the fourth moment formula it can be shown that for *Gaussian* processes the covariance of the direct estimate at different frequencies is given by

$$\mathrm{Cov}\{\hat{S}_D(\omega+\zeta), \hat{S}_D(\omega-\zeta)\} = \left|\frac{1}{2\pi}\int S(\zeta-\xi)\tilde{D}^*(\xi+\omega)\tilde{D}(\xi-\omega)d\xi\right|^2 + \left|\frac{1}{2\pi}\int S(\omega-\xi)\tilde{D}^*(\xi+\zeta)\tilde{D}(\xi-\zeta)\,d\xi\right|^2 \quad (17)$$

In this equation the first term is large only in the neighborhood of the origin ($\omega = 0$) while the second term is a convolution which, for $\zeta = 0$, equals $\mathrm{E}\{\hat{S}_D(\omega)\}^2$. If, on the other hand, we set $\zeta = \Delta/2$, the second term gives the covariance of estimates with a frequency separation of Δ in the vicinity of ω. It is helpful to view the direct estimate, $\hat{S}_D(\omega)$, as a nonstationary time series with a *known* covariance structure and in regions where the spectrum is locally white as a *stationary* series. As with other stationary series the second-order properties of the direct estimate in such regions are described by an autocorrelation function, which for unit spectrum is given by

$$\Lambda(\Delta) = |\tilde{D}^*\tilde{D}^*|^2 \quad (18)$$

Figure 5 shows the autocorrelation functions of the different direct spectrum estimates as a function of frequency separation, and again the local properties of the prolate tapers are striking by comparison to the very poor properties of the other estimates. It should be noted, however, that for the 4π window at the usual frequency mesh spacing of $1/2T$ the correlation between estimates is 0.9077 so that, as shown in Fig. 4, the distribution of estimates at this spacing is quite different than it is for independent estimates.

It is frequently more convenient to work with the Fourier transform, Ξ_D, of this autocorrelation which we call the *antespectrum* of the estimator. Thus Ξ_D is defined by

$$\Xi_D(Q) = D^2(Q)*D^2(Q) \quad (19)$$

and is the *spectrum of the spectrum estimate*, $S_D(\omega)$. The antespectrum is a function of *quefrency*, Q, which is a lag or time-like variable and its Fourier transform is the autocorrelation function, Λ, of *the spectrum estimate* expressed as a function of ordinary frequency separation.

Defining a smoothed direct estimate $S_{D,W}(\omega)$ as

$$\hat{S}_{D,W}(\omega) = \frac{1}{2\pi}\int \hat{S}_D(\omega-\zeta)W(\zeta)\,d\zeta \quad (20)$$

in which the weight, W, is usually considered to be symmetric, positive, and integrating to 1. Since the spectral window of the direct estimate, $S_D(\omega)$, is $|\tilde{D}(\omega)|^2$ the spectral window of the smoothed estimate is clearly

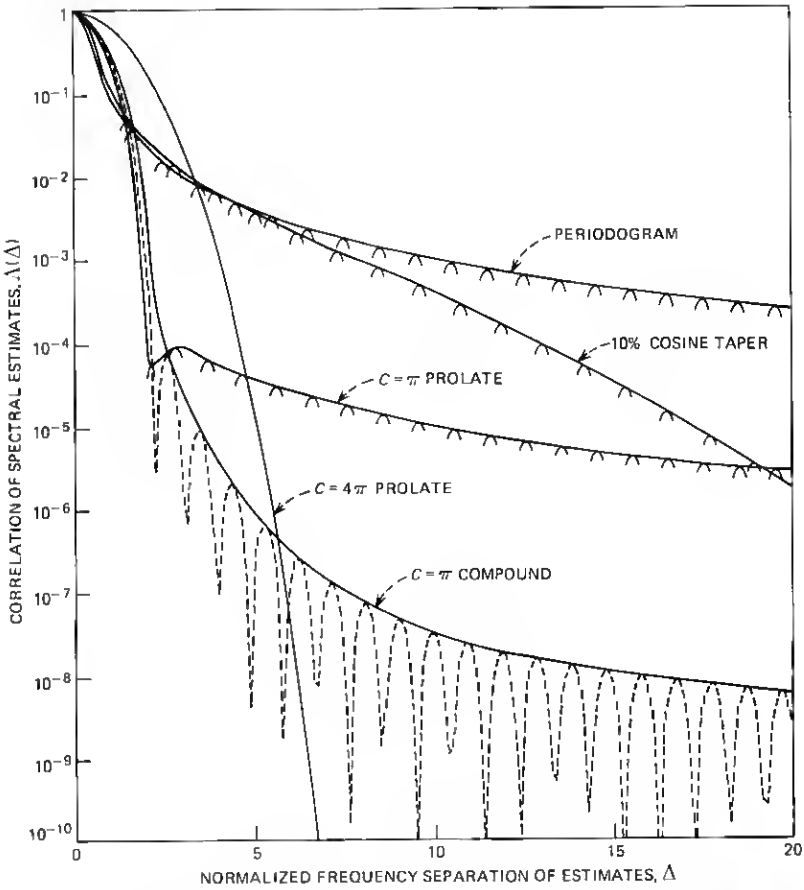

Fig. 5—Envelopes of autocorrelation functions of direct spectrum estimates.

the convolution $|\tilde{D}|^{2} * W$ but the variance does not correspond to the usual interpretation of a spectral window in the literature on indirect estimates. Using the above definitions the influence of a smoothing operation, or *lifter*, may be described in the quefrency domain as a linear filter so that the antespectrum, $\Xi_{D,W}$, or the *spectrum of the smoothed spectrum estimate* is the product of the antespectrum, Ξ_D, and the power transfer function of the lifter. The variance of the smoothed spectrum estimate is the integral, over quefrency, of its antespectrum so that the estimate $\hat{S}_{D,W}$ will have an approximately χ^2 distribution with

$$\nu_{D,W} = 2\left[\int_{-T}^{T} |\tilde{W}(Q)|^{2} \Xi_D(Q)\, dQ\right]^{-1} \tag{21}$$

equivalent degrees of freedom. For direct estimates the antespectrum

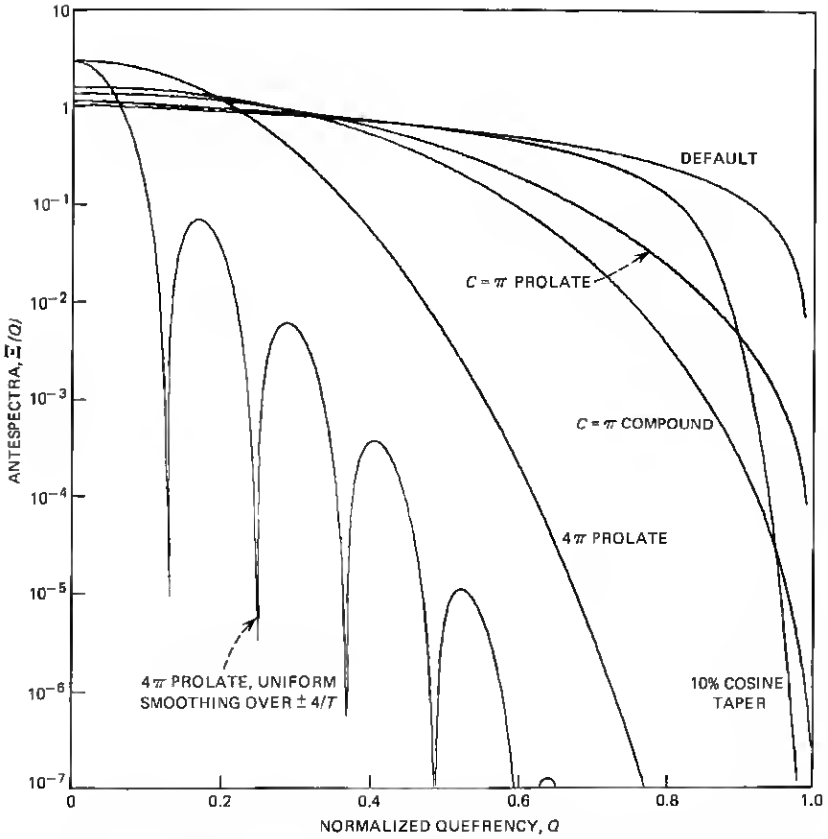

Fig. 6—Antespectra corresponding to various data windows (the antespectrum is the spectrum of the spectrum estimate.

is symmetric with a global maximum at zero quefrency. From eq. (21) it is clear that for liftering to be effective $|\tilde{W}|^2$ should be small when Ξ_D is large.

As indicated above the choice of weights is a complex subject which depends to a large extent on the intended application with perhaps the best linear smoothers obtained by modifying the technique described by Papoulis[44] to account for the data window. When this is done the Sturm-Liouville equation

$$\frac{\partial}{\partial t}(D^2(t)y') + \lambda y = 0 \tag{22}$$

is obtained corresponding to his eq. (22) and can be solved by standard techniques.

Figure 6 shows the antespectra corresponding to the various data windows. From these curves it is apparent that estimates using the 4π

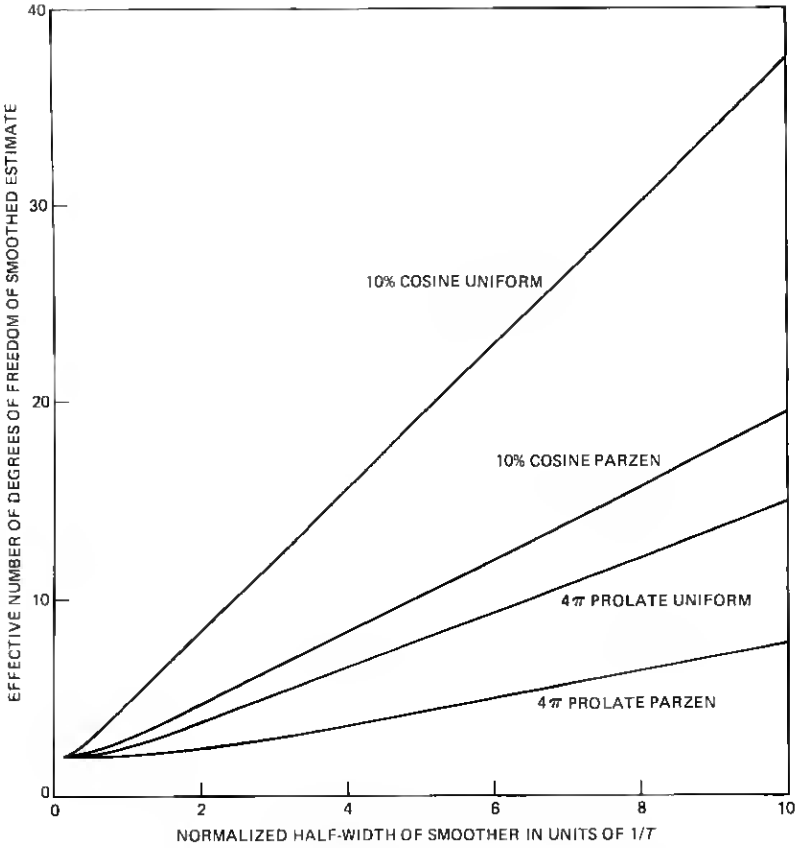

Fig. 7—Effectiveness of smoothing direct estimates.

taper have more of their variance at low frequencies than the other estimates. The bottom curve in this figure shows the antespectrum of a 4π estimate smoothed with uniform weights over a bandwidth of $\pm 4/T$ which, when integrated, gives only 6.7 equivalent degrees of freedom. Figure 7 is a plot of the equivalent degrees of freedom resulting when direct estimates are smoothed. Two smoothers are used: simple moving averages which are useful for calibration purposes and the modified Parzen weights (Cleveland and Parzen[46]). *From these curves it is obvious that the correlation induced in the raw spectrum estimate by the data window can result in significantly fewer degrees of freedom than expected on the basis of uncorrelated spectrum estimates.* This effect is particularly noticeable with the 4π taper where, when bias considerations are excluded, the asymptotic efficiency is only 36 percent when only a single direct estimate is computed. As will be seen in the next section, the use of overlapped subsets results in variance efficiency.

3.3 The Welch technique and optimum subset spacings

An alternative to smoothing across frequency is Welch's method[11] (see also Cooley *et al.*[47]) in which the data is divided into several overlapping subsets, direct estimates computed on each subset, and the results combined. The individual subset estimates have the usual statistical properties of direct estimates but when used jointly one must also account for the correlation between subsets. For the same reasons given in the previous section the effectiveness of this averaging must he accurately determined. Clearly spacing the subsets too close results in computational inefficiency while if they are spaced too far apart the procedure is statistically inefficient.

Consider two direct estimates of spectra $\hat{S}_1$ and $\hat{S}_2$ made on the domains $(0, T)$ and $(b, b + T)$ respectively. For Gaussian processes the covariance between these estimates is given by

$$\mathrm{Cov}[\hat{S}_1(\omega), \hat{S}_2(\omega)|b] = \left|\frac{1}{2\pi}\int S(\zeta)e^{i\zeta b}\bar{D}(\omega - \zeta)\bar{D}^*(\omega + \zeta)\,d\zeta\right|^2 + \left|\frac{1}{2\pi}\int S(\zeta)e^{i\zeta b}|\bar{D}(\omega - \zeta)|^2\,d\zeta\right|^2 \quad (23)$$

The first term of this expression is large near the frequency origin but elsewhere the second term dominates. Since $\bar{D}$ is an entire function it is clear that the covariance hetween the two estimates is governed primarily by the characteristics of the actual spectrum S, in the vicinity of ω. In particular spectra having very narrow resonances or discontinuous characteristics will result in the subsets being correlated for large values of the offset b. The effect of this correlation is that averaging the different subset estimates does not give the usual reduction of variance so that the autoregressive model is unstable when only a few subsets are available. When the correlation between suhsets is low the distribution of the average of k subsets is nearly χ^2_{2k}.

When the spectrum is locally smooth estimates of this type depend, in addition to the data window, on the two parameters T and b. The length of the individual subsets depends primarily on the fine structure of the process and will be discussed in Section V. The relative spacing of subsets, however, depends largely on the choice of the data window and in general there is an optimum spacing. Under the usual approximation that the true spectrum is locally constant or linear and that we are interested in frequencies away from the origin, eq. (23) simplifies, and the correlation between subsets becomes the square of the equivalent lag window, $L_D(b)$.

As a measure of effectiveness of this procedure, assume that sufficient data is available to compute k subsets. Standardizing the local spectral level to 1, the variance of the averaged estimate is

$$V_k(b) = \frac{1}{k} + \frac{2}{k}\sum_{s=1}^{k-1}\left(1 - \frac{s}{k}\right) L_D^2(sb) \tag{24}$$

We now consider the effect of adding sufficient new data to compute $k + 1$ subsets and, by analogy with Fisher information, we measure the relative gain in information by

$$\Delta I_k(b) = \frac{1}{b}\left[\frac{1}{V_{k+1}(b)} - \frac{1}{V_k(b)}\right] \tag{25}$$

As k becomes large ΔI_k rapidly approaches the limit

$$\Delta I_\infty(b) = \frac{1}{b}\frac{1}{1 + 2\sum_{s=1}^{[T/b]} L_D^2(sb)} \tag{26}$$

This function is plotted in Fig. 8 for the different windows discussed earlier. When the subsets are spaced very closely relative to their length, no information is "missed" by falling between adjacent subsets, but on the other hand the subsets are highly correlated with each other so that the addition of a subset does not decrease the variance very much. For the 4π prolate window this situation remains true until the spacing between subsets becomes about 0.25 to 0.30 of their length, after which the information recovery becomes rapidly less efficient. Because the computational burden rapidly increases as the offset is decreased, a subset spacing of about 0.29 of the subset length is used. For the less concentrated windows this effect is less important. It should also be noted that the higher information recovery of the 4π prolate window evident here is consistent with the fact that it has a broader frequency response, so that, apart from bias considerations, the overall efficiencies of the techniques are similar. When bias considerations are included the efficiency of the prolate window is much higher.

IV. DATA PREPARATION

Assuming that aliasing and noise effects have been properly kept at a minimum there are usually two steps of data preparation necessary in time series work. The first is the elimination of *gross* errors and the second is the removal of *deterministic* mean value functions.

Gross errors are inevitable in very large data sets, see Hampel,[48] and experience has shown that the WT4 project is no exception to this rule.

Errors which are large and easily visible are best removed at an early stage in the processing. A simple strategy which works for both large errors and missing values is as follows:

(*i*) Data points in serious error are tagged, either on the basis of vi-

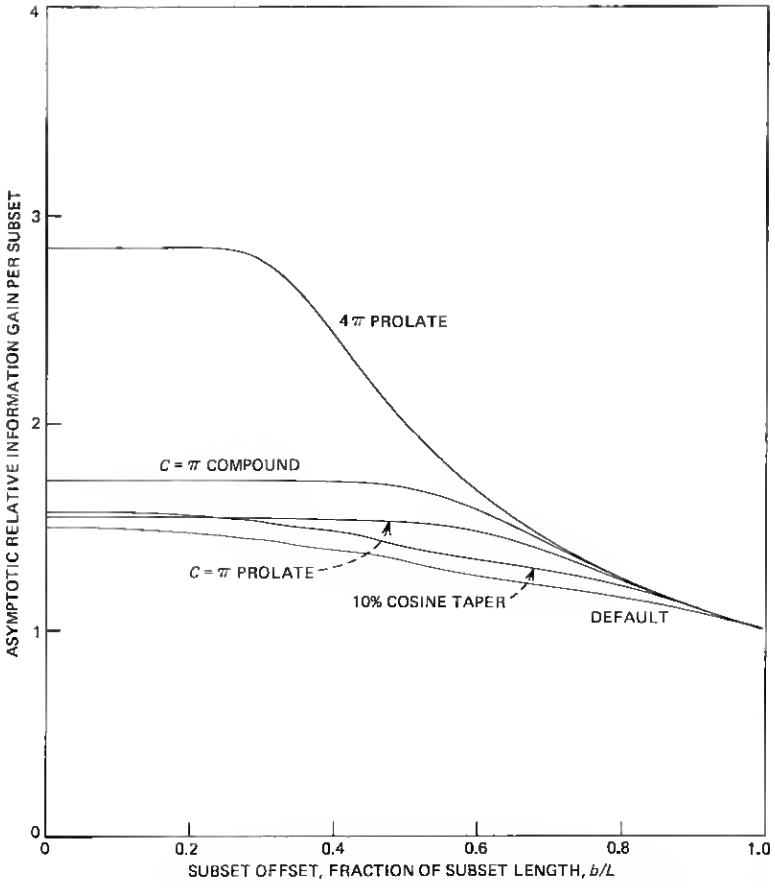

Fig. 8—Asymptotic relative information gain as a function of subset base offset.

sual examination or automatically on the basis of built in validity checks.†

(*ii*) Predictors and interpolators were generated either from untagged data or, if similar data sets were available, from the autocorrelations of these sets.

(*iii*) When the error is isolated it is replaced by interpolation. If the errors could not be considered isolated those adjacent to the longest stretch of good data were corrected first by prediction from the good section. After all the tagged points had been initially corrected by prediction, the corrections were recomputed by interpolation using the initial correction as a basis for the interpolations.

† Records of long range mouse data were coded to provide indications of hardware malfunction.

However, not all errors in time series data are obvious from a plot. In situations where the spectrum covers many decades an error may be insignificant on the scale of the *process* variance but catastrophic compared to the *innovations* variance. At the data preparation stage these errors are neither easily detectable nor troublesome and so their correction is deferred to the prewhitening part of the process where they are effectively eliminated by the robust filtering process.

The other stage of data preparation is the removal of deterministic mean value functions from the data. A simple example of a series with a nonconstant mean value function is given by the axis curvature of a waveguide following a planned route bend.

The usual approach in time series analysis with problems of this type is to remove the "trend" using orthogonal polynomial regression techniques. This approach has proved unsatisfactory primarily because such a high-degree polynomial is required to approximate the mean value function that the residuals bear little resemblance to the stochastic part of the process.

A method of removing trends in data which has proved generally effective is based on the use of polynomial B-splines. A B-spline of *order* k is a piecewise continuous polynomial of degree $k - 1$ defined by an array of *knots*, some of which may be multiple. The continuity properties of these functions are controlled by the knots; the spline is discontinuous at a knot of multiplicity k, has a discontinuous derivative at knots of multiplicity $k - 1$, and so on. At simple knots, or knots of multiplicity 1, the spline has $k - 2$ continuous derivatives. Details of the theory of B-splines are contained in a paper by Curry and Schoenberg,[49] a recent paper by de Boor[50] describes computational aspects, and Horowitz[51] discusses the characteristics of splines with equispaced simple knots in terms of their frequency domain characteristics.

Figure 9 shows a plot of the measured elevation of a waveguide line and an approximate mean value function generated through the use of B-splines. By choosing a spline with few knots, indicated on the figure, a simple fit to the gross topology of the run is obtained so that the "roughness" of the installation is readily apparent.

A second example of the use of polynomial spline mean value functions is shown in Fig. 10, which is a plot of the vertical output from a measurement of axis curvature on a waveguide tube supported on Airy point supports (see Fox *et al.*[52]). In this case most of the indicated curvature is a result of the tube sagging under its own weight and this effect is readily calculable and is shown by the dashed line. As a check that this removal is not distorting the spectrum of the actual distortions in the tube the ratio, an F statistic, of the average of 10 estimates of the spectrum of the detrended vertical curvature to the spectrum of the hori-

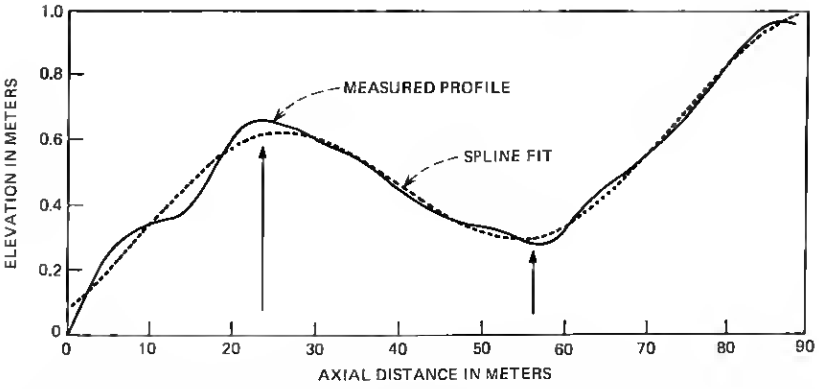

Fig. 9—Example of a B-spline fit to waveguide elevation knots. The spline is order 4 with knots of multiplicity 4 at −0.1 and 90 meters, simple knots at 23.5 and 56.5 meters.

zontal curvature was computed. No significant differences could be detected between the two sets of spectra.

V. PILOT SPECTRUM ESTIMATE

The actual process used to generate the pilot spectrum estimate is a combination of the two smoothing approaches described in Section III. The use of subsets allows the test for stationarity and, because this test is more sensitive when applied to smoothed data, a logical step is to smooth the subset estimates individually. However for the stationarity test to be effective it is necessary that the different subset estimates be essentially uncorrelated at any given frequency. This requirement results in the base offset between adjacent subsets being more than about 57 percent of the subset length, which is larger than is desirable for the most effective use of the data from an information recovery viewpoint. The obvious solution is to compute the subsets with the 29 percent offset mentioned above and use every other subset in the stationarity test.

A further advantage of the use of subsets is that a significant improvement in the accuracy of the pilot estimate can often be obtained by combining the different subset estimates in a robust manner instead of by the usual arithmetic average. Denoting the ordered subset estimates by $\vec{S}_j(\omega)$ with $\vec{S}_1(\omega) \leqslant \vec{S}_2(\omega) \leqslant \cdots \leqslant \vec{S}_{k'}(\omega)$, a robust estimate $\overleftrightarrow{S}(\omega)$ may be formed as

$$\overleftrightarrow{S}(\omega) = \sum_{j=1}^{k'} \Theta_j \vec{S}_j(\omega) \tag{27}$$

where the weights, $\{\Theta_j\}$, which depend on k', are chosen so that $\overleftrightarrow{S}$ is a minimum variance unbiased estimate of S. General techniques for forming such estimates are given in Lloyd[53] and the specific means and

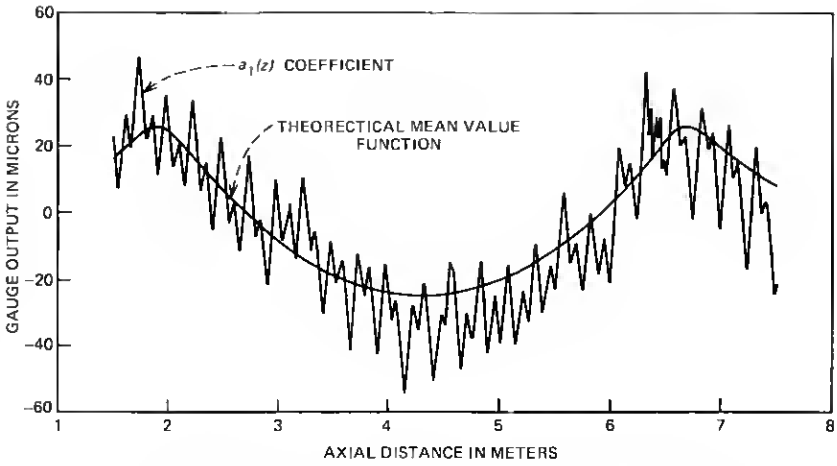

Fig. 10—Example of a spline mean value function. Data represent vertical curvature output from a rotating-head mouse.

covariances of the order statistics for gamma distributions required are given in Sarhan and Greenberg[54] and Prescott.[55]

For the unsmoothed subset estimates the means and covariances are particularly simple and the weights are given explicitly by

$$\Theta_j = \frac{1}{k'} \qquad j < k' \tag{28}$$

$$\Theta_{k'} = \frac{k + 1 - k'}{k'}$$

and the variance of $\vec{S}(\omega) = \mathrm{E}\{\hat{S}(\omega)\}^2/k'$ so that the efficiency, relative to an uncensored estimate, is just k'/k. It is shown by Mehrata and Nanda[56] that this estimate is maximum likelihood. This procedure is most effective for eliminating the effects of the occasional gross outlier missed in the data preparation stage but, unlike the robust filter algorithm, is ineffective against numerous small outliers.

As mentioned earlier, the length of the individual subsets is dependent on the fine structure of the spectrum to be estimated. A simple method of estimating this length (which within fairly broad bounds is not critical since the final estimate is primarily responsible for fine structure) is to compute a moving average representation of tbe process. For this purpose the Wiener canonical spectral factorization approach is ideally suited and, if in eq. (42) below, the sign of the summation is reversed and the expression Fourier-transformed, a moving average representation‡

‡ This moving average representation is the minimum delay causal nonrecursive (transverse) filter generating the observed process from white noise. The convolution of the moving average with itself gives the autocovariance function of the process.

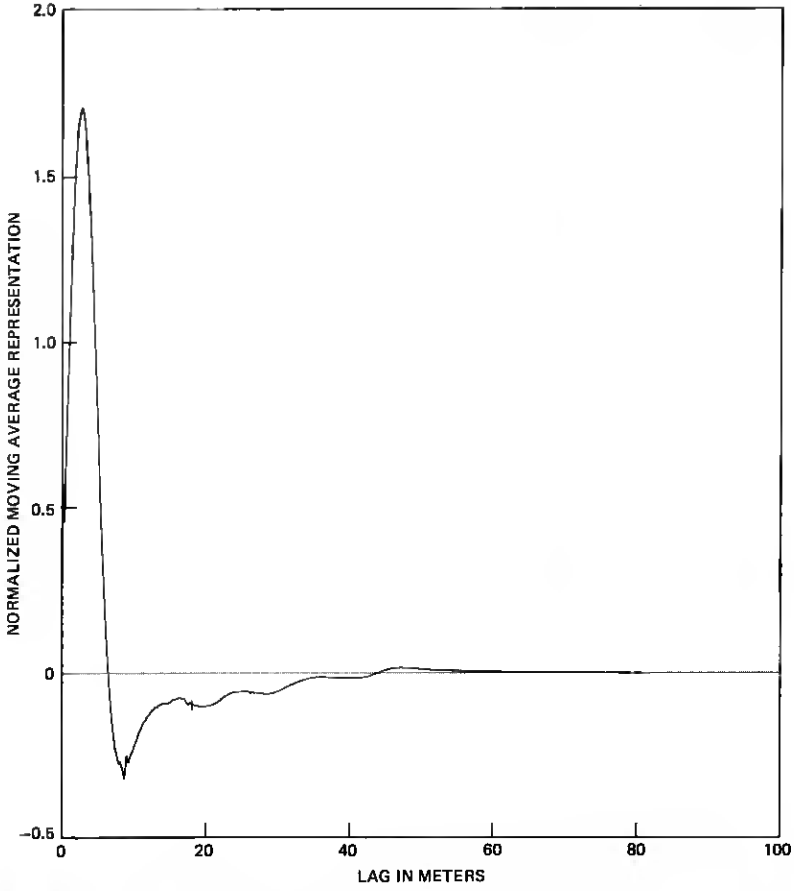

Fig. 11—Canonical moving average representation of vertical curvature gauge output. Representation based on the average of 102 450-meter subsets.

is obtained instead of an autoregression. Figure 11 shows such a model for the vertical gauge output for the Netcong field trial data and it is apparent that most of the weight is concentrated within a 60-meter range.† To allow for the heavy tapering effect of the 4π prolate window a subset length of 160 meters was used.

VI. CONSTRUCTION OF AUTOREGRESSIVE MODELS

The basic reason for computing a pilot spectrum estimate is to permit the design of an accurate prewhitening filter and the subsequent use of

† The discontinuities visible in this plot near 9 and 18 meters are due to the couplings but due to the randomized guide lengths this effect is rapidly suppressed with increasing separation.

a high-resolution spectrum estimation technique on the filtered data. Prewhitening filters are subject to several constraints; for example their transfer function must have no zeroes in the frequency range of interest [see eq. (61)], their design must be readily automated, they must have finite impulse response, they must be numerically well conditioned, and they must be absolutely stable. The prediction error filter satisfies these requirements.

Moreover, since the prediction error filter is causal and depends on the canonical autoregressive representation of a stationary time series it has the further major advantage that it may be readily robustified as described in Section VII. The use of the prediction error filter is therefore conditioned on one's ability to estimate the parameters of an autoregressive model of the process and this problem is the subject of the present section. Current work on autoregressive modeling uses two distinct approaches; direct solution of the Yule-Walker equations such as described by Pagano,[57] Ulrych and Bishop,[58] Makhoul,[59] and spectral factorization as described in Bhansali.[60,61] Following a brief review of these two approaches a composite technique is described which exploits features of both. The section concludes by considering three alternative methods of computing prewhitening filters.

In the *autoregressive* representation of a discrete time process the value, x_t, of the series at time t is given by the sum of a regression on the past values of the series and an independent random component, ξ_t,

$$x_t = \xi_t + \sum_{j=1}^{p} \alpha_j x_{t-j} \tag{29}$$

An equivalent description is to regard the regression on the past of the series as a *prediction* of the value of the process at time t so that the random component, ξ_t, represents the "new" information or *innovations* of the process. The length of the predictor or order of the regression is denoted by p which may be infinite. Such processes and questions related to them are discussed extensively in the literature; see for example Hannan,[62] Koopmans,[24] Box and Jenkins,[63] or Doob.[28] The papers by Kailath,[64] Kailath and Frost[65] are also relevant to these problems.

6.1 Yule-Walker equations and the Lavinson algorithm

The basic equations determining the autoregressive coefficients are derived by minimizing the one-step prediction variance

$$\sigma_p^2 = \mathrm{E}\left\{\left(x_t - \sum_{j=1}^{p} \alpha_j^{(p)} x_{t-j}\right)^2\right\} \tag{30}$$

with respect to α_j for $j = 1, 2 \cdots p$ and are known as the *Yule-Walker equations*

$$\rho_k = \sum_{j=1}^{p} \alpha_j^{(p)} \rho_{|j-k|} \qquad k=1, 2, \cdots p \tag{31}$$

In these equations ρ_k is the autocorrelation function of the $\{x\}$ process at lag k. These equations are not only *linear* in the α's but also *Toeplitz*, so that the matrix elements depend only on their distance from the main diagonal and for real series the $p \times p$ matrix has only p distinct elements. Since these equations are linear, they may be solved using standard techniques such as the QR algorithm (see Dahlquist *et al.*[66]). However, because of their special structure, special procedures are available for their solution which require only p^2 operations instead of the p^3 required with general linear equation techniques. Also, because fewer operations are required, roundoff errors are reduced and the faster algorithms can be more accurate.

Generally these fast algorithms are similar in structure to the recursive solution of the Yule-Walker equations discovered by Levinson.[67] One convenient and numerically stable variant is due to Durbin,[69] which in the notation of Ramsey,[70] is initiated using

$$\phi_1 = \alpha_1^{(1)} = \rho_1 \tag{32}$$

$$\sigma_1^2 = 1 - \phi_1^2 \tag{33}$$

and continued for $k = 1, 2, \cdots, p-1$ by

$$\phi_{k+1} = \alpha_{k+1}^{(k+1)} = \left\{ \rho_{k+1} - \sum_{j=1}^{k} \alpha_j^{(k)} \rho_{k+1-j} \right\} \Big/ \sigma_k^2 \tag{34}$$

$$\alpha_j^{(k+1)} = \alpha_j^{(k)} - \phi_{k+1} \alpha_{k+1-j}^{(k)} \qquad j = 1, 2 \cdots, k \tag{35}$$

$$\sigma_{k+1}^2 = \sigma_k^2 (1 - \phi_{k+1}^2) \tag{36}$$

In these equations the $\alpha_j^{(k)}$'s are the autoregressive or prediction coefficients for k step prediction, the ϕ sequence is known as the *partial autocorrelation function*, and σ_k^2 is the k step relative prediction error. In the original Levinson algorithm the expansion

$$\sigma_k^2 = 1 - \sum_{j=1}^{k} \alpha_j^{(k)} \rho_j \tag{37}$$

obtained by substituting eq. (31) into (30) was used in place of eq. (36). Analytically these equations are identical but the latter is both slower and also has much poorer numerical properties than Durbin's form.

6.2 Spectral factorization

One drawback of tbe Toeplitz matrix formulation is that it does not provide much insight into the actual minimization process and it is helpful to rewrite the equations in terms of a *prediction error filter* where

we define

$$\alpha_0^{(p)} = -1 \tag{38}$$

and the negative error sequence

$$z_i = -(x_i - \hat{x}_i) \tag{39}$$

$$= \sum_{k=0}^{p} \alpha_k^{(p)} x_{i-k}$$

Note that the $\{z\}$ sequence is a result of a linear causal convolution, or filtering operation, applied to the $\{x\}$ sequence. The transfer function of this filter is

$$A^{(p)}(\omega) = \sum_{k=0}^{p} \alpha_k^{(p)} e^{-i\omega k} \tag{40}$$

so that the spectrum of $\{z\}$ is $S_x(\omega)|A^{(p)}(\omega)|^2$ with the corresponding variance

$$\sigma_p^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega)|A^{(p)}(\omega)|^2 \, d\omega \tag{41}$$

where $S_x(\omega)$ is the spectrum of the $\{x\}$ process. As $p \to \infty$ the spectrum of the error sequence approaches a constant so that the problem is to choose the causal filter in such a way that $|A|^2$ is small whenever S_x is large. For A a trigonometric polynomial of degree p the problem has been completely solved by Szegö[71] and the recursion formulae for the orthogonal polynomials obtained are essentially similar to those above.

An alternative solution is provided by Wiener's[68] canonical spectral factorization where the filter transfer function A is represented as

$$A(\omega) = -\exp\left\{-\sum_{k=1}^{\infty} c_k e^{-i\omega k}\right\} \tag{42}$$

so that the variance of $\{z\}$ is given by

$$\sigma_z^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) e^{-2\sum_{k=1}^{\infty} c_k \cos\omega k} \, d\omega \tag{43}$$

Direct minimization of this expression as a function of the c_k's is impractical due to the complexity of the resulting equations. Wiener's approach is to identify the c_k with the Fourier series coefficients of $\ln S_x$, that is

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos \omega k \, \ln \{S_x(\omega)\} \, d\omega \tag{44}$$

The sequence c_k is referred to as the *cepstrum*.

It is important to notice that the series in eq. (42) does not include a c_0 term because the constraint imposed by eq. (38) implies that c_0 defines

the minimum. This is most easily seen by substituting the Fourier series representation (44) into eq. (41) with the result

$$\sigma_I^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\{S_x(\omega)\}\, d\omega\right] \tag{45}$$

σ_I^2 is the *innovations* variance of the process. For this procedure to be valid the spectral distribution function must be absolutely continuous which implies that autoregressive representations are invalid for process containing periodic components. The procedure is formal and as mentioned in Wiener and Masani,[72] the sense in which it converges is unknown.

D. Preston[73] has observed that in practice these convergence problems may be avoided by evaluating the formula (Rozanov[74])

$$A(\omega) = \lim_{\mu\to 1} \frac{-1}{\sigma_I} \exp\left\{-\frac{1}{4\pi}\int_{-\pi}^{\pi} \ln S(\lambda)\frac{e^{-i\lambda} + \mu e^{-i\omega}}{e^{-i\lambda} - \mu e^{-i\omega}} d\lambda\right\} \tag{46}$$

inside it's radius of convergence, *ie* for $\mu < 1$, ratber than on the radius of convergence as does the Wiener approach. With this modification one obtains

$$A_\mu(\omega) = -\exp\left\{-\sum_{k=1}^{\infty} c_k \mu^k e^{-i\omega k}\right\} \tag{47}$$

so that the coefficients α_j may be computed by Fourier transforming $A_\mu(\omega)$ and dividing by μ^j.

These two techniques have been described as if the actual spectrum were known. When applied to an estimate of the spectrum, things are more complex and neither technique has a clear advantage over the other. The disadvantage of the first approach is that it works explicitly with the autocorrelation function. The range of spectra common in waveguide work is so large that use of the autocorrelation function is numerically undesirable in that information corresponding to the lower parts of the spectrum may be lost due to numerical roundoff errors. The second method is numerically stable but produces a filter with a very long impulse response which reproduces all the details of the spectrum on which it is based, including those due to sampling. Since the robust filter algorithm works in the time domain the shortest autoregressive model which retains the statistically significant features of the spectrum is desirable.

Cleveland[75] and Bartholomew[76] have described several sources of error in prediction problems. Of these the most critical appears to be a result of sampling variability in the spectrum estimate. As an example consider the estimate of innovations variance, $\hat{\sigma}_I^2$ obtained by using the pilot spectrum estimate, $\overline{S}$, in place of the spectrum, S_x, in eq. (45). This estimate is described by Davis and Jones[77] except that their bias correction

is not used in steps involving the model formulation. (The bias correction is used to set the scale of the residuals for the robust filter algorithm.) Grenander and Rosenblatt[18] give a formula for prediction error in the case when the predictor is based on an estimate of spectrum, $\hat{S}$, rather than on the true spectrum, S, of the process

$$\hat{\sigma}_f^2 = \sigma_f^2 \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\frac{\hat{S}(\omega)}{S(\omega)}\,d\omega\right\}\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{S(\omega)}{\hat{S}(\omega)}\,d\omega \tag{48}$$

Periodogram estimates are distributed as χ_2^2 so that $\mathrm{E}\{\hat{S}^{-1}(\omega)\} = \infty$. While this result is based on the Wiener spectral factorization method and so applies to prediction using the entire past, it appears to give a good indication of the behavior of autoregressive fits even for relatively compact predictors. Further information on the effects of smoothing on the estimated innovations variance is available in Jones.[78]

6.3 Reduced factorization

A method which exploits many of the advantages of both of the preceding approaches without having the fatal flaws of either is to reduce the result of spectral factorization. In the reduced factorization approach one begins by creating a long autoregressive model using Wiener's spectral factorization method described above and then converts it to a shorter representation using the Levinson recursion formulae. In this reduction the key equation is (35) which, by combining the updates for $\alpha_j^{(k)}$ and $\alpha_{k+1-j}^{(k)}$, may be written backwards. When written for use in a downwards recursion this formula becomes

$$\alpha_j^{(k)} = \frac{\alpha_j^{(k+1)} + \phi_{k+1}\alpha_{k+1-j}^{(k+1)}}{1 - \phi_{k+1}^2} \tag{49}$$

Similarily the k-step prediction variance, σ_k^2 may be obtained from σ_{k+1}^2 by using eq. (36) backwards starting from the estimate of innovations variance given by eq. (45).

The major disadvantage of the reduced factorization technique is that it is somewhat slower than either of the standard techniques individually. Of these the Levinson recursion is the faster: it requires only a single Fourier transform to convert the pilot spectrum to a sample autocorrelation function and then p^2 operations for the actual solution. In practice it is necessary to "search" for the correct order of the autoregression. Since this search is never carried past $p_{\max} = \sqrt{T}$ the total computation time is $\sim T \ln T$. Since spectral factorization requires three Fourier transform operations, its speed is comparable with that of the Levinson technique. Reduced factorization requires an additional T^2 operations and is therefore considerably slower when very large data subsets are being used.

As mentioned earlier the most serious flaw with the Levinson approach is a result of roundoff errors *in the Fourier transform* used to convert the pilot spectrum estimate to autocovariances and is only serious when the range of the pilot spectrum estimate is large. Roundoff characteristics of fast Fourier transform algorithms are well understood (see Kaneko and Liu[79]) and consequently the characteristics of the pilot spectrum estimate relative to the computer precision may be used to select the "best" procedure: when the range of the pilot estimate is low the Levinson-Durbin algorithm is used, but in cases where the range is large reduced factorization is preferred.

With either approach the order of the autoregressive representation, p, has been chosen as the value of τ for which Parzen's[16] criterion

$$P(\tau) = 1 - \frac{\hat{\sigma}_f^2}{\hat{\sigma}_\tau^2} + \frac{\tau}{T} \tag{50}$$

attains its minimum. Within reasonable bounds the actual order selected is not critical as the autoregressive model is used as a prewhitening filter and not as a spectrum estimate. (The function $\hat{\sigma}_p^2/|A^{(p)}(\omega)|^2$ is known as an *autoregressive spectrum estimate.* See Akaike,[80] Gersch and Sharpe.[81]) Berk[82] gives conditions on the order, p, for obtaining a consistent model of the process.

The actual method used to determine the prediction error filter is a combination of the two methods discussed in Sections 6.2 and 6.3 as shown in the flow diagram, Fig. 12. In its general form this spectrum estimation technique is an iterative process and intermediate estimates are used to update the pilot estimate of spectrum and the prediction error filter. In cases when iteration is used it is stopped when the estimated innovations variance stabilizes.

6.4 Alternatives

Since the sequences of steps which is being used here to generate an autoregressive model is by no means obvious it is worthwhile to briefly examine the alternatives. The obvious technique of eliminating the pilot spectrum estimation and transformation to autocorrelations procedure and estimating the sample autocorrelations directly is not done because of the high bias, discussed in Section IV, of this estimate.

The second possibility is to form the pilot estimate of spectra and then design a conventional digital filter for the prewhitening operation. Details of this approach using Gegenbauer filters are given in Thomson.[83] The drawback is that such filters are incompatible with the robust filter algorithm.

A third alternative is to directly estimate the partial correlations using Burg's[84] algorithm. In this approach an autoregressive model is estimated

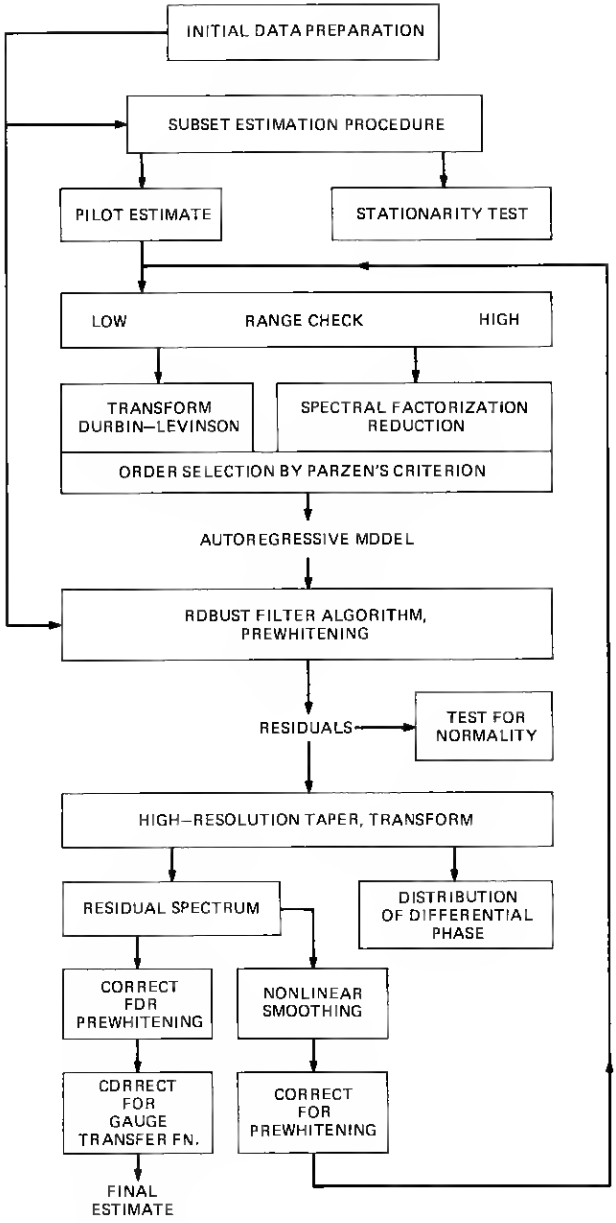

Fig. 12—Iterative estimation procedure.

by minimizing the sum of the forward and reverse prediction errors

$$e^{2(p)} = \sum_{n=1}^{T-p} \left(x_n - \sum_{k=1}^{p} \alpha_k^{(p)} x_{n+k} \right)^2 + \sum_{n=p+1}^{T} \left(x_n - \sum_{k=1}^{p} \alpha_k^{(p)} x_{n-k} \right)^2 \quad (51)$$

successively for $p = 1, 2 \cdots$ under the constraint that the covariance matrix is Toeplitz so that the autoregressive coefficients, $\alpha_k^{(p)}$, are updated using eq. (35). This method, also referred to as the "maximum entropy" approach, gives the partial autocorrelations directly and constrains them to be less than 1 in magnitude.

Limited Monte-Carlo studies indicate that spectra based on autoregressive representations obtained in this way have exceptionally high variance whenever the order, p, of the autoregression is carried far enough to reveal details of the spectrum. Other problems with the Burg algorithm as a *spectrum analysis technique* are described in Chen and Stegun.[85] As a *prewhitening algorithm* it has been used on individual tubes with reasonable success. These "partial Burg" routines have been effective in situations where a *low* order autoregressive representation is adequate for the prewhitening filter, the range of the spectrum is low, and very compact code is required.

VII. ROBUST FILTERING AND PREWHITENING

One of the most useful data analysis tools developed during the course of this work is the *robust filter algorithm*. This is a nonlinear technique designed to eliminate the effects of occasional "outliers" in the data from the final spectrum estimate, where, as mentioned earlier, outliers are measured on the scale of the innovations process. As an example of the magnitude of this problem, the typical output of a tubing curvature gauge is about 15 microns rms whereas the scale of the innovations process is about 0.5 microns. This is considerably smaller than the size of typical dust particles and, since this gauge operates primarily in a tubing mill, the probability of some dust particles being measured is quite high and the need for robust filtering is evident.

The robust filter algorithm differs from linear filtering in that most of the data passes through the "filter" without modification and only those points which are basically unpredictable from past values of the series are changed. The characteristics of the filtering algorithm are controlled by providing (*i*) an autoregressive model of the process, (*ii*) an estimate of the innovations variance, and (*iii*) an influence function. In practice the autoregressive model and innovations variance must be estimated from the data and it has been found that the algorithm works well even with surprisingly inacurate models. Further details and examples on this procedure are available in Kleiner *et al.*[86] The steps of this procedure, as it is currently implemented, are listed below. Section 7.2 summarizes results (see Kleiner *et al.*[87] for details) relevant to the choice of influence function, and in Section 7.3 an example of the action on contaminated data is given. Further information on robust procedures is available in Huber[88] and Hampel.[48]

7.1 The robust filter algorithm

We assume that the observations, $\{y\}$, consist of the process of interest, $\{x\}$, plus occasional outliers, $\{v\}$

$$y_k = x_k + v_k \tag{52}$$

Based on this contaminated data the robust filter algorithm produces an estimate, $\{\hat{x}\}$, of the "core" process by the following steps:

(*i*) A prediction, $\tilde{x}_n$, is made from the filtered sequence using the autoregressive coefficients obtained by the methods discussed in Section VI.

$$\tilde{x}_n = \sum_{k=1}^{p} \alpha_k^{(p)} \hat{x}_{n-k} \tag{53}$$

(*ii*) A weight is defined which depends on the difference, $y_n - \tilde{x}_n$, between the actual observation, y_n, and the prediction. This difference is normalized by the scale of the innovations process, σ_p. (The scale is the square root of the prediction variance estimate, $\hat{\sigma}_p^2$, with the bias correction given in Davis and Jones.[77])

$$w_n = W\left(\frac{y_n - \tilde{x}_n}{\sigma_p}\right) \tag{54}$$

In the applications described here W is an even function with $W(0) = 1$ and $W(\infty) = 0$. When multiple errors are encountered the scale, σ_p, used in this formula is replaced by an approximation of the k-step prediction variance.

(*iii*) The output of the robust filter algorithm is an estimate of the core process, $\hat{x}_n$, formed by the weighted average of observation and prediction

$$\hat{x}_n = w_n y_n + (1 - w_n)\tilde{x}_n \tag{55}$$

The effect of this procedure is to leave the data unmodified where the prediction errors are small and to replace the data with its prediction at points where the prediction errors are gross. The action taken when the prediction errors are near the expected extreme for the given sample size depends on the weight function which will be discussed below. In spectrum estimation applications the desired output is usually not the filtered sequence but rather the prewhitening residuals

$$z_n = \hat{x}_n - \tilde{x}_n \tag{56}$$

The z_n may be described in terms of an influence function (see Hampel[89])

$$\psi(e) = eW(e) \tag{57}$$

applied to the relative prediction error, $(y_n - \tilde{x}_n)/\sigma_p$, but the notation is deceptive in that it deemphasizes the fact that the weighting procedure also influences the prediction for subsequent steps.

To use this algorithm the filtered sequence must be initialized on the first p points. When the data is only slightly contaminated the raw data has been used to start the process but when the contamination is more severe special precautions must be taken.

The detailed behavior of the algorithm depends on the choice of weight function, and this represents a compromise between rejecting valid outliers of the innovations and accepting the occasional erroneous data point. Considerable information is available on the choice and characteristics of influence functions for robust estimates of location (see Andrews *et al.*[90]), but this information is of limited utility in time series applications since in location estimates there is no concern with frequency response characteristics. It must be remembered that this operation is *nonlinear* and that nonlinear operations on time series generally change the spectrum in complex ways. Because of this the weight or influence function must be chosen in such a way that the spectral content due to the induced nonlinearities is much less than that due to the presence of errors in the data.

Several different weights have been used. Of these the best found to date is a result of motivation by the extreme value distribution for distributions of exponential type (see Kendall and Stuart[91]) and is defined by

$$W(u) = \exp\{-e^{u_0(|u|-u_0)}\} \tag{58}$$

in which

$$u_0 = \Phi^{-1}\left(1 - \frac{1}{N}\right) \tag{59}$$

Φ being the normal cumulative distribution function and N the sample size. This influence function, shown in Fig. 13 for $N = 1000$, is very linear in the center and, at about $\pm 3\sigma$, decreases rapidly to zero.

7.2 Spectral distortions resulting from robust filtering

In its most general form, use of the robust filter algorithm is alternated with the model formation process as shown in Fig. 12. In this iterative mode the output from the filter is used to generate a better autoregressive model which is used to filter the data and so on. This kind of iterative procedure has been used for some difficult data sets and was found to converge to a stable estimate of spectrum very rapidly. Typically two or three iterations are required on short series (for example, some distortions in individual tubes) where the range of the spectrum is very large

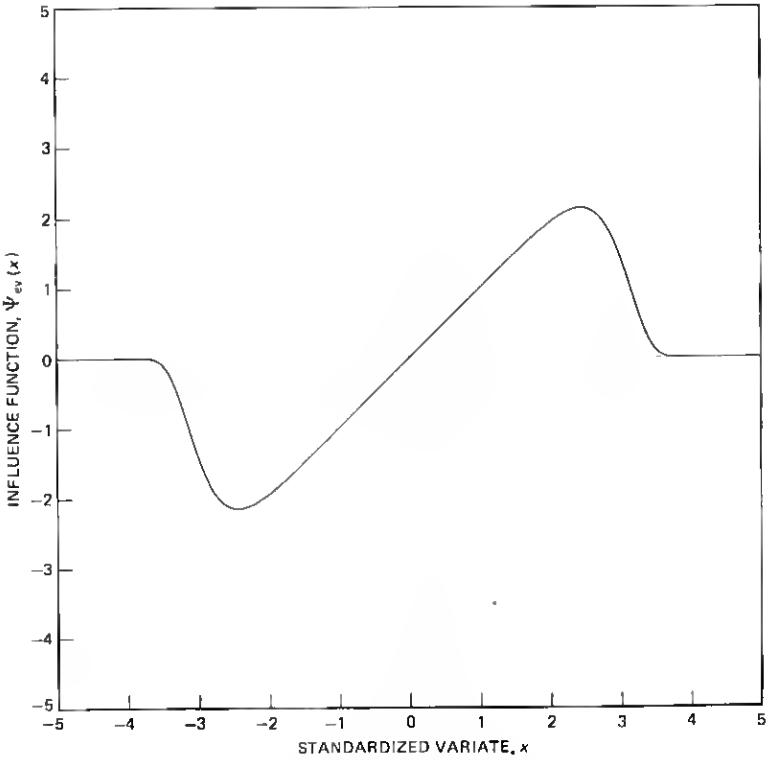

Fig. 13—Extreme value influence function.

and the outliers are small relative to the scale of the process but large compared to the scale of the innovations. With very large data sets, such as those from complete mode filter sections of the field trial (which average 80,000 data points), a single iteration has been used and found satisfactory.

If one assumes that this iterative process has converged, it is possible to describe the distortions introduced into the spectral density estimate. At convergence the autoregressive parameters, $\hat{\alpha}_k$, describe the estimated process, $\{\hat{x}\}$, and are solutions of the Yule-Walker equations based on the estimated process. Then by computing the expectation of $\hat{x}_{n-k}$ with $\hat{x}_n$ using the representation eq. (55) for the latter, it is found that the $\hat{\alpha}_k$'s also are solutions of a set of *robust* Yule-Walker equations

$$\mathrm{E}\left\{\hat{x}_{n-k}\psi\left[\frac{y_n - \sum_{j=1}^{p}\hat{\alpha}_j\hat{x}_{n-j}}{\sigma_p}\right]\right\} = 0 \qquad k = 1, \cdots, p \tag{60}$$

An alternative viewpoint is to regard the algorithm as the solution of minimizing a nonquadratic loss function of $y_n - \bar{x}_n$ with respect to the $\hat{a}_k$'s and it can be shown that the solution to this problem also yields the robust Yule-Walker equations provided that the influence function is such that the error sequences, $\partial \hat{x}_n / \partial \hat{a}_j$ are small. This is a reasonable requirement: small changes in the process specification should not result in large changes in the filter output. The satisfaction of this condition depends on the choice of influence function, ψ. It can be shown that the scale of the error sequences depends on $1 - \psi'$ so that influence functions having very high curvature in regions where the probability density function of the innovations process is large result in larger errors than influence functions which are more linear in such regions. The most important property of the algorithm, however, is that, for reasonable influence functions, the effect of the nonlinearities on the spectrum estimate, is proportional to the spectrum so that the net effect is a slight downwards bias. The scale of the bias factor is $E\{\xi\psi(\xi)\}$ and, for the dominant error terms, is independent of frequency.

7.3 Action of the robust filter on contaminated data

The intent of the robust filter algorithm is to reduce the effects of outliers and erroneous data from the final estimate of spectra. Since the choice of influence function is to some extent distribution dependent, it is also of interest to observe the effect of this algorithm in a direct manner. It is also interesting to check to what extent a normal assumption on the basic data is warranted. Since the high serial correlations existing in most time series in the physical sciences make the usual tests for goodness-of-fit to a given distribution inapplicable this must be done cautiously. A very conservative approach is to find some lag, τ_0, such that the autocorrelations at multiples of this lag are small and test samples taken at this spacing for normality. Since the spacing required to obtain uncorrelated data may be large, this approach is rather inefficient. An alternative is to consider the residuals from the prewhitening operation. Since these residuals are generally very small, usually only a few times the quantization level, this method is very sensitive to outliers and measurement errors. Figure 14 shows a Q-Q plot of the residuals from a *linear* prewhitening operation and it is clear that the apparent distribution has very heavy tails. If the actual residuals are plotted as a time series, Fig. 15, it is clear that at least part of the long-tailed characteristics are a simple consequence of the fact that in linear prewhitening each outlier in the original series is converted into $p + 1$ outliers in the residual series. In Fig. 16 a Q-Q plot of the residuals from the robust prewhitening algorithm is given and the contrast is striking. In this case the residuals are quite close to normal and in agreement with tests made on other sections of the line.

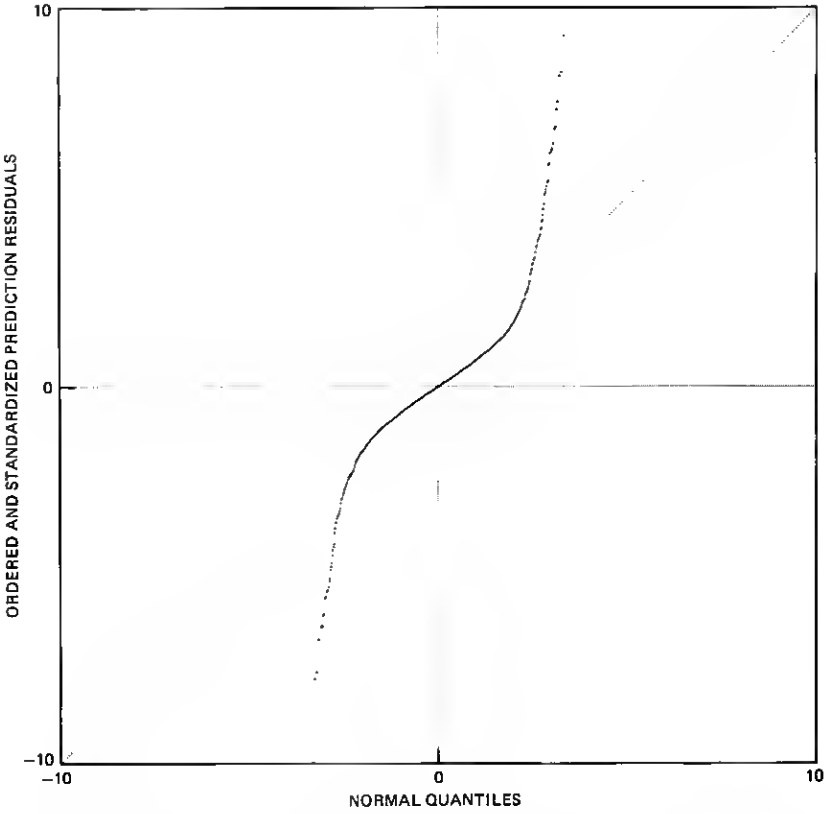

Fig. 14—WT4 field evaluation test horizontal curvature gauge output. Residuals from a linear prediction error filter.

VIII. FINAL ESTIMATE OF SPECTRUM

Prewhitening converts the data from a highly correlated into an almost uncorrelated form whose spectrum has a low dynamic range. Estimates of such spectra are best made with windows which have high frequency resolution and do not need the extreme sidelobe suppression used for the pilot estimate and the Tukey spliced cosine window has been used for most such applications.

The final estimate is intended primarily to extract *details* of the process: consequently the data is *not* split into subsets and the estimate is not smoothed by liftering. In cases where "smootbing" is done it has been by the nonlinear methods discussed in Section 2.7. These techniques might be described as "inverse influence" in that individual points are lumped into a moving average *except* when they are outliers in which case they are used instead of the average. This procedure is a useful aid for spotting peaks and other low level features in the spectrum.

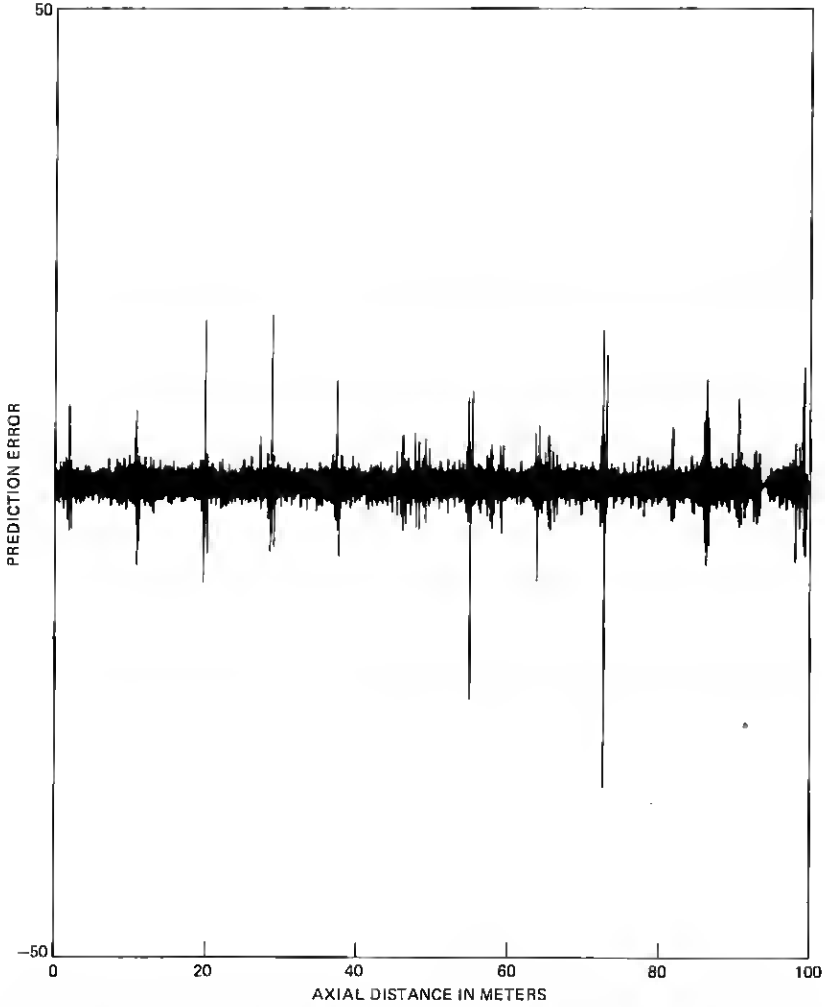

Fig. 15—WT4 field evaluation test horizontal curvature gauge output. Residuals from a linear prediction error filter.

In the final estimate of spectrum it is necessary to correct for the prewhitening operation so that the result is expressed as the ratio

$$\hat{S}(\omega) = \frac{S_z(\omega)}{\left| \sum_{k=0}^{P} \alpha_k e^{-i\omega k} \right|^2} \tag{61}$$

in which $S_z(\omega)$ is a direct estimate of the spectrum of the prewhitened residuals [eq. (56)], and the denominator is the power transfer function of the prediction error filter defined in eq. (39).

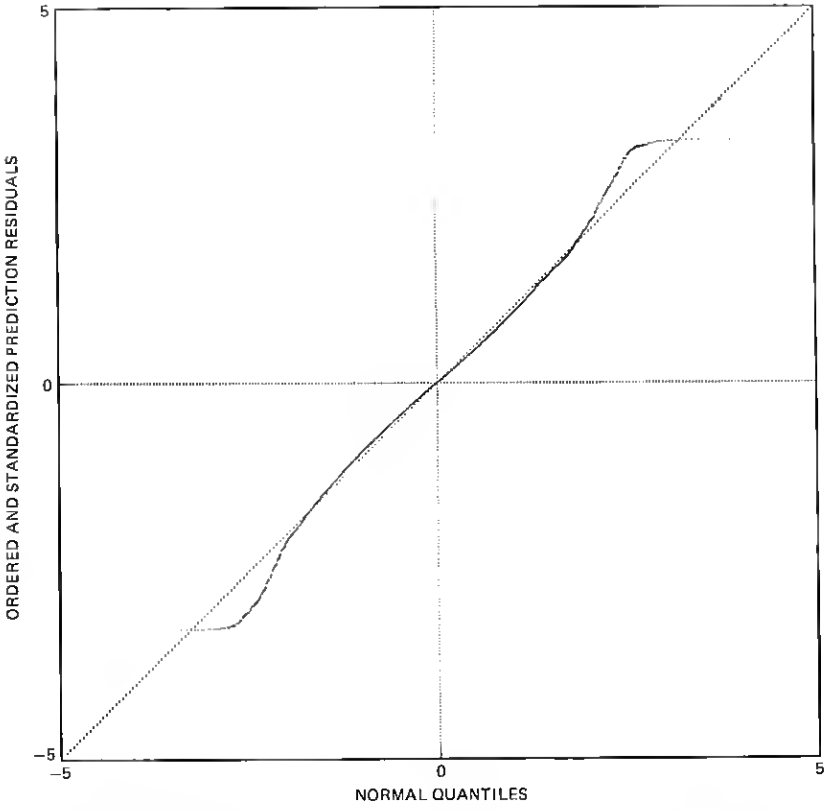

Fig. 16—WT4 field evaluation trial curvature gauge output. Prediction residuals from the robust filter algorithm.

The validity of this form depends on the assumption of the independence of the filter and the residuals and is discussed briefly in Grenander and Rosenblatt.[18] With the prediction error filter this is a reasonable assumption in that the filter depends on the partial autocorrelation functions up to lag p while any residual structure is primarily the contribution of the partial correlations for higher lags. This assumption is also supported by Whittle's[15] observation that the information matrix splits into one part describing the structure of the process and a second part describing the innovations sequence.

IV. CONCLUSIONS OF PART I

A technique for estimating the power spectral density function of a stationary time series has been described which is robust, accurate, and computationally straightforward. Part II of this paper will give examples of its use and comparisons with standard techniques.

APPENDIX A

Formulae for prolate spheroidal data windows

A convenient expansion of the prolate spheroidal wave function data windows is given in Flammer[38] Section 3.2. This expansion is a power series in terms of

$$U = (1 - x)(1 + x)$$

Computationally it is advantageous to rewrite the power series using Horner's rule and for $c = 4\pi$ the expansion is:

$$D_{00}(4\pi, x) = \sqrt{\frac{2\Delta x}{.508125548147497T}}\,((((((((((((((((((((+2.6197747176990866d-11)U+2.9812025862125737d-10)U+3.0793023552299688d-09)U+2.8727486379692354d-08)U+2.4073904863499725d-07)U+1.8011359410323110d-06)U+1.1948784162527709d-05)U+6.9746276641509466d-05)U+3.5507361197109845d-04)U+1.5607376779150113d-03)U+5.8542015072142441d-03)U+1.8482388295519675d-02)U+4.8315671140720506d-02)U+1.0252816895203814d-01)U+1.7233583271499150d-01)U+2.2242525852102708d-01)U+2.1163435697968192d-01)U+1.4041394473085307d-01)U+5.9923940532892353d-02)U+1.4476509897632850d-02)U+1.5672417352380246d-03)U+4.2904633140034110d-05)$$

The expansion for the higher resolution window with $c = \pi$ is:

$$D_{00}(\pi, x) = \sqrt{\frac{2\Delta x}{T}}\,(((((((((+5.3476939016920851d-11)U+2.2654256220146656d-09)U+7.8075102004229667d-08)U+2.1373409644281953d-06)U+4.5094847544714943d-05)U+7.0498957221483167d-04)U+7.7412693304064753d-03)U+5.5280627452077586d-02)U+2.2753754228751827d-01)U+4.3433904277546202d-01)U+2.2902051859068017d-01)$$

In the forms given here botb functions have been normalized for use as

data windows. In this application x takes on values

$$x_t = \frac{2t-1}{T} - 1; \qquad t = 1, 2, \ldots, T$$

REFERENCES

1. J. C. Anderson *et al.*, B.S.T.J., to be published.
2. R. A. Fisher, *Statistical Methods and Scientific Inference* (3d ed), Hafner Press, 1973.
3. M. Arato, "On the Sufficient Statistics for Stationary Gaussian Processes," Theory Probab. Appl., *6* (1961), pp. 199–201.
4. P. Whittle, "Estimation and Information in Stationary Time Series," Arkiv För Matermatik, *2* (1953), pp. 423–434.
5. D. J. Thomson, "Spectral Analysis of Sbort Series," thesis, Polytechnic Institute of Brooklyn, 1971.
6. M. Loève, *Probability Theory*, D. Van Nostrand, 1963.
7. U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, Univ. of Cal. Press, 1958.
8. H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—II," B.S.T.J., *40*, No. 1 (January 1961), pp. 65–84.
9. K. O. Dzhaparidze and A. M. Yaglom, "Asymtotically Efficient Estimation of the Spectrum Parameters of Stationary Stochastic Processes," Proc. Prague Symp. on Asymptotic Statistics, *1*, Prague: Charles Univ. Press, 1974.
10. E. A. Aronson, "Fast Fourier Integration of Piecewise Polynomial Functions," Proc. IEEE, *57* (1969), pp. 691–692.
11. P. D. Welcb, "The Use of the Fast Fourier Transform for Estimation of Spectra: A Method Based on Time Averaging Over Sbort, Modified Periodograms," IEEE Trans. Audio Electroacoust., *AU-15* (1967), pp. 70–74.
12. D. J. Thomson, "A Test for Stationarity," 1977.
13. T. Kailath, "A View of Three Decades of Linear Filtering Theory," IEEE Trans., *IT-20* (1974), pp. 146–180.
14. J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, *63* (1975), pp. 563–580.
15. P. Whittle, *Prediction and Regulation by Linear Least-Squares Methods*, D. Van Nostrand, 1963.
16. E. Parzen, "Some Recent Advances in Time Series Modelling," IEEE Trans., *AC-19* (1974), pp. 723–730.
17. D. C. Rife and G. A. Vincent, "Use of the Discrete Fourier Transform in the Measurement of Frequencies and Levels of Tones," B.S.T.J., *49*, No. 2 (February 1970), pp. 197–228.
18. U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, New York: Wiley, 1975.
19. R. B. Blackman and J. W. Tukey, "Tbe Measurement of Power Spectra," B.S.T.J., *37*, Nos. 1 and 3 (January and March 1958), pp. 185–282, 485–569. (Reprinted by Dover.)
20. R. H. Jones, "Spectral Estimates and Their Distributions," Skandinavisk Aktuatietidskrift, *45* (1962), pp. 39–69, 135–153.
21. R. H. Jones, "A Reappraisal of the Periodogram in Spectral Analysis," Technometrics, *7* (1965), pp. 531–542.
22. R. H. Jones, "Spectrum Estimation with Missing Observations," Ann. Inst. Stat. Math., *23* (1971), pp. 387–398.
23. J. W. Tukey, "An Introduction to the Calculations of Numerical Spectrum Analysis," *Spectral Analysis of Time Series*, B. Harris, ed., New York: Wiley, 1967.
24. L. H. Koopmans, *The Spectral Analysis of Time Series*, Academic Press, 1974.
25. D. Brillinger, *Times Series, Data Analysis and Theory*, Holt, Rinehart & Winston, 1975.
26. V. F. Pisarenko, "On the Estimation of Spectra by Means of Non-Linear Functions of the Covariance Matrix," Geophysical J., Royal Astronomical Soc., *28* (1972), pp. 511–531.
27. J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," Proc. IEEE, *57* (1969), pp. 1408–1418.
28. J. L. Doob, *Stochastic Processes*, New York: Wiley, 1953.

29. D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—I," B.S.T.J., *40*, No. 1 (January 1961), pp. 43–64.
30. H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—III," B.S.T.J., *41*, No. 4 (July 1962), pp. 1295–1336.
31. D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—IV," B.S.T.J., *43*, No. 9 (November 1964), pp. 3009–3057.
32. D. Slepian and E. Sonnenblick, "Eigenvalues Associated with Prolate Spheroidal Wave Functions of Zero Order," B.S.T.J., *44*, No. 8 (Octoher 1965), pp. 1745–1759.
33. D. Slepian, "Some Asymptotic Expansions for Prolate Spheroidal Wave Functions," J. Math. Physics, *44* (1965), pp. 99–140.
34. D. J. Thomson, M. F. Rohhins, C. G. MacLennan, and L. J. Lanzerotti, "Spectral and Windowing Techniques in Power Spectral Analysis of Geomagnetic Data," Physics of the Earth and Planetary Interiors, *12* (1976), pp. 217–231.
35. J. F. Kaiser, "Nonrecursive Digital Filter Design Using the I_0 – sinh Window Function," IEEE Inter. Symp. Circuits & Systems Proc. (1974), pp. 20–23.
36. A. Eberhard, "An Optimal Discrete Window for the Calculation of Power Spectra," IEEE Trans., *AU-21* (1973), pp. 37–43.
37. E. Parzen, "Mathematical Considerations in the Estimation of Spectra," Technometrics, *3* (1961), pp. 167–190.
38. C. Flammer, *Spheroidal Wave Functions*, Stanford Univ. Press, 1967.
39. R. P. Bogert, M. J. Healy, and J. W. Tukey, "The Frequency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking," *Time Series Analysis*, M. Rosenhlatt, ed., pp. 209–243, New York: Wiley, 1962.
40. R. A. Fisher, "Tests of Significance in Harmonic Analysis," Proc. Royal Soc. London, *125 A* (1929), pp. 54–59.
41. M. S. Bartlett, "Some Remarks on the Analysis of Time Series," Biometrika, *54* (1967), pp. 25–38.
42. K. S. Miller, R. I. Bernstein, and L. E. Blumenson "Generalized Rayleigh Processes," Quart. Jour. Math., *16* (1958), pp. 137–145.
43. E. Parzen, "On Consistent Estimates of the Spectrum of a Stationary Time Series," Ann. Math. Stat., *28* (1957), pp. 329–348.
44. A. Papoulis, "Minimum-Bias Windows for High-Resolution Spectral Estimates," IEEE Trans., *IT-19* (1973), pp. 9–12.
45. U. Grenander, H. O. Pollak, and D. Slepian, "The Distrihution of Quadratic Forms in Normal Variates: A Small Sample Theory with Applications to Spectral Analysis," J. SIAM, *7* (1959), pp. 374–401.
46. W. S. Cleveland and E. Parzen, "The Estimation of Coherence, Frequency Response, and Envelope Delay," Technometrics, *17* (1975), pp. 167–172.
47. J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "The Application of the Fast Fourier Transform Algorithm to the Estimation of Spectra and Cross-Spectra," *Computer Processing in Communications*, Polytechnic Institute of Brooklyn Microwave Research Institute Symposia Series, *19* (1969), pp. 5–20.
48. F. R. Hampel, "Rohust Estimation: A Condensed Partial Survey," Z. Wahrscheinlichkeitstheorie verw., *27* (1973), pp. 87–104.
49. H. B. Curry and I. J. Schoenherg, "On Polya Frequency Functions IV; The Fundamental Spline Functions and Their Limits," Jour. Analyse Math., *17* (1966), pp. 71–107.
50. C. de Boor, "Package for Calculating with B-Splines," SIAM Jour. Numer. Analysis, *14* (1977), pp. 441–472.
51. L. L. Horowitz, "The Effects of Spline Interpolation on Power Spectra Density," IEEE Trans., *ASSP-22* (1974), pp. 22–27.
52. P. E. Fox, S. Harris, and D. J. Thomson, "Mechanical Gauging Techniques," B.S.T.J., to be puhlished.
53. E. H. Lloyd, "Least Squares Estimation of Location and Scale Parameters Using Order Statistics," Biometrika, *39* (1952), pp. 88–95.
54. A. E. Sarhan and B. G. Greenberg, *Contributions to Order Statistics*, New York: Wiley, 1962.
55. P. Prescott, "Variances and Covariances of Order Statistics from the Gamma Distrihution," Biometrika, *61* (1974), pp. 607–613.
56. K. G. Mehrata and P. Nanda, "Unhiased Estimation of Parameters hy Order Statistics in the Case of Censored Samples," Biometrika, *61* (1974), pp. 601–606.
57. M. Pagano, "An Algorithm for Fitting Autoregressive Schemes," Jour. Royal Stat. Soc., C *21* (1972), pp. 274–281.
58. T. J. Ulrych and T. N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Rev. Geophys. Space Phys., *13* (1975), pp. 183–200.

59. J. Makhoul, "Spectral Linear Prediction: Properties and Applications," IEEE Trans., *ASSP-23* (1975), pp. 283–296.
60. R. J. Bhansali, "A Monte-Carlo Comparison of the Regression Method and the Spectral Methods of Prediction," Jour. Amer. Stat. Assoc., *68* (1973), pp. 621–625.
61. R. J. Bhansali, "Asymptotic Properties of the Wiener-Kolmogorov Predictor," Jour. Royal Stat. Soc., 1974, pp. 61–73.
62. E. J. Hannan, *Multiple Time Series*, New York: Wiley, 1970.
63. G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day, 1970.
64. T. Kailath, "An Innovations Approach to Least-Squares Estimation—Part I: Linear Filtering in Additive White Noise," IEEE Trans., *AC-13* (1968), pp. 645–655.
65. T. Kailath and P. Frost, "An Innovations Approach to Least-Squares Estimation—Part II: Linear Smoothing in Additive White Noise," IEEE Trans., *AC-13* (1968), pp. 655–660.
66. G. Dahlquist, Å. Björk, and N. Anderson, *Numerical Methods*, New York: Prentice-Hall, 1974.
67. N. Levinson, "The Wiener RMS Error Criterion in Filter Design and Prediction," Jour. Math. Physics, *25* (1947), pp. 261–278. (Reprinted as Appendix B of Wiener.[68])
68. N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, M.I.T. Press, 1949.
69. J. Durbin, *Distribution Theory for Tests Based on the Sample Distribution Function*, SIAM, 1973.
70. F. L. Ramsey, "Characterization of the Partial Autocorrelation Function," Ann. Stat., *2* (1974), pp. 1296–1301.
71. G. Szegö, *Orthogonal Polynomials*, third ed., American Matb. Society, 1967.
72. N. Wiener and P. Masani, "The Prediction Theory of Multivariate Stochastic Processes, II—The Linear Predictor," Acta Mathematica, *99* (1958), pp. 93–137.
73. D. B. Preston, "private communication," 1977.
74. Y. A. Rozanov, *Stationary Stochastic Processes*, Holden Day, 1967.
75. W. S. Cleveland, "Fitting Time Series Models for Prediction," Technometrics, *13* (1971), pp. 713–723.
76. D. J. Bartholomew, "Errors of Prediction for Markov Chain Models," Jour. Royal Stat. Soc., *B 37* (1975), pp. 444–456.
77. H. T. Davis and R. H. Jones, "Estimation of the Innovations Variance of a Stationary Time Series," Jour. Amer. Stat. Assoc., *63* (1968), pp. 141–149.
78. R. H. Jones, "Estimation of the Innovation Generalized Variance of a Multivariate Stationary Time Series," Jour. Amer. Stat. Assoc., *71* (1976), pp. 386–388.
79. T. Kaneko and B. Liu, "Accumulation of Round-off Error in Fast Fourier Transforms," J. Assoc. Comp. Mach., *17* (1970), pp. 637–654.
80. H. Akaike, "Power Spectrum Estimation Through Autoregressive Model Fitting," Ann. Inst. Stat. Math., *21* (1969), pp. 407–419.
81. W. Gersch and D. R. Sharpe, "Estimation of Power Spectra with Finite-Order Autoregressive Models," IEEE Trans., *AC-18* (1973), pp. 367–369.
82. K. N. Berk, "Consistent Autoregressive Spectral Estimates," Ann. Stat., *2* (1974), pp. 489–502.
83. D. J. Thomson, "Generation of Gegenbauer Prewhitening Filters by Iterative Fast Fourier Transforming," *Computer Processing in Communications*, Polytechnic Institute of Brooklyn Press, 1969.
84. J. P. Burg, "Maximum Entropy Spectral Analysis," thesis, Stanford Univ., 1975.
85. W. Y. Ghen and G. R. Stegen, "Experiments with Maximum Entropy Power Spectra of Sinusoids," J. Geophysical Res., *79* (1974), pp. 3019–3022.
86. B. Kleiner, R. D. Martin, and D. J. Thomson, "Three Approaches Towards Making Power Spectra Less Vulnerable to Outliers," Bus. & Econ. Stat. Sect., Proc. Amer. Stat. Assoc. (1976), pp. 386–391.
87. B. Kleiner, R. D. Martin, and D. J. Thomson, "Robust Estimates of Spectra," in preparation.
88. P. J. Huber, "Robust Statistics: a Review," Ann. Math. Stat., *43* (1972), pp. 1041–1067.
89. F. R. Hampel, "The Influence Curve and its Role in Robust Estimation," Jour. Amer. Stat. Assoc., *69* (1974), pp. 383–393.
90. D. F. Andrews *et al.*, *Robust Estimates of Location*, Princeton Univ. Press, 1972.
91. M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, 1*, New York: Hafner, 1963.